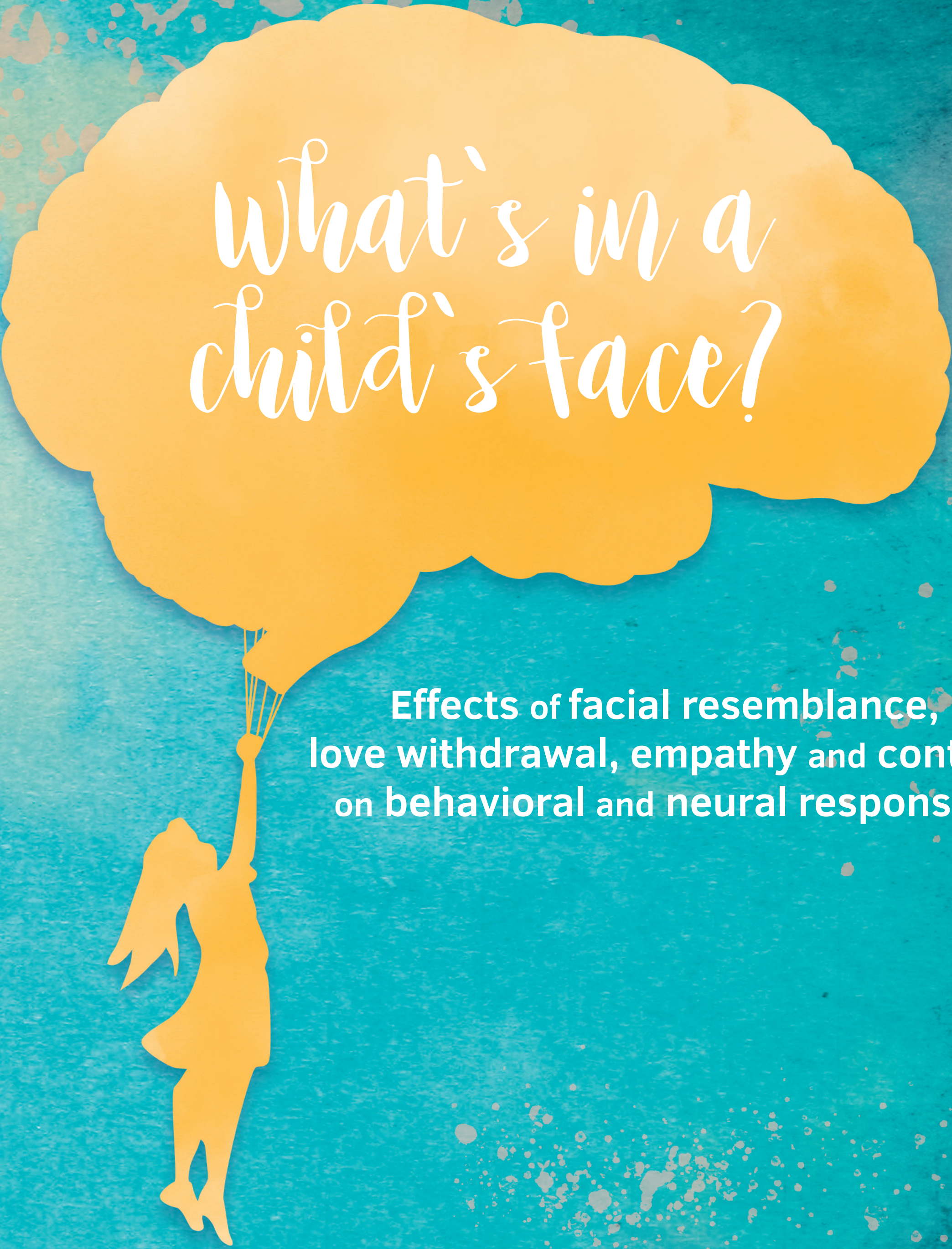
What's in a child's face? : effects of facial resemblance, love withdrawal, empathy and context on behavioral and neural responses

Heckendorf, E.

Citation

Heckendorf, E. (2018, April 17). *What's in a child's face? : effects of facial resemblance, love withdrawal, empathy and context on behavioral and neural responses*. Retrieved from https://hdl.handle.net/1887/61482

Author: Heckendorf, Esther
Title: What's in a child's face? : effects of facial resemblance, love withdrawal, empathy and context on behavioral and neural responses
Date: 2018-04-17



what`s in a child`s face?

Effects of facial resemblance, love withdrawal, empathy and context on behavioral and neural responses

Esther Heckendorf

What’s in a child’s face?

Effects of facial resemblance, love withdrawal, empathy and context on behavioral and neural responses.

Printed by Mostert & Van Onderen, 2018

Cover design by Evelien Jagtman

ISBN: 978-94-90858-54-4

What's in a child's face?

Effects of facial resemblance, love withdrawal, empathy and context on behavioral and neural responses.

PROEFSCHRIFT

ter verkrijging van de graad van Doctor aan de Universiteit Leiden
op gezag van de Rector Magnificus prof. mr. C. J. J. M. Stolker,
volgens besluit van het college voor promoties
te verdedigen op dinsdag 17 april 2018
klokke 11:15 uur
door

Esther Heckendorf

geboren te Andernach, Duitsland , in 1989

Promotores:
Prof. dr. M.J. Bakermans-Kranenburg
Prof. dr. M. H. van IJzendoorn

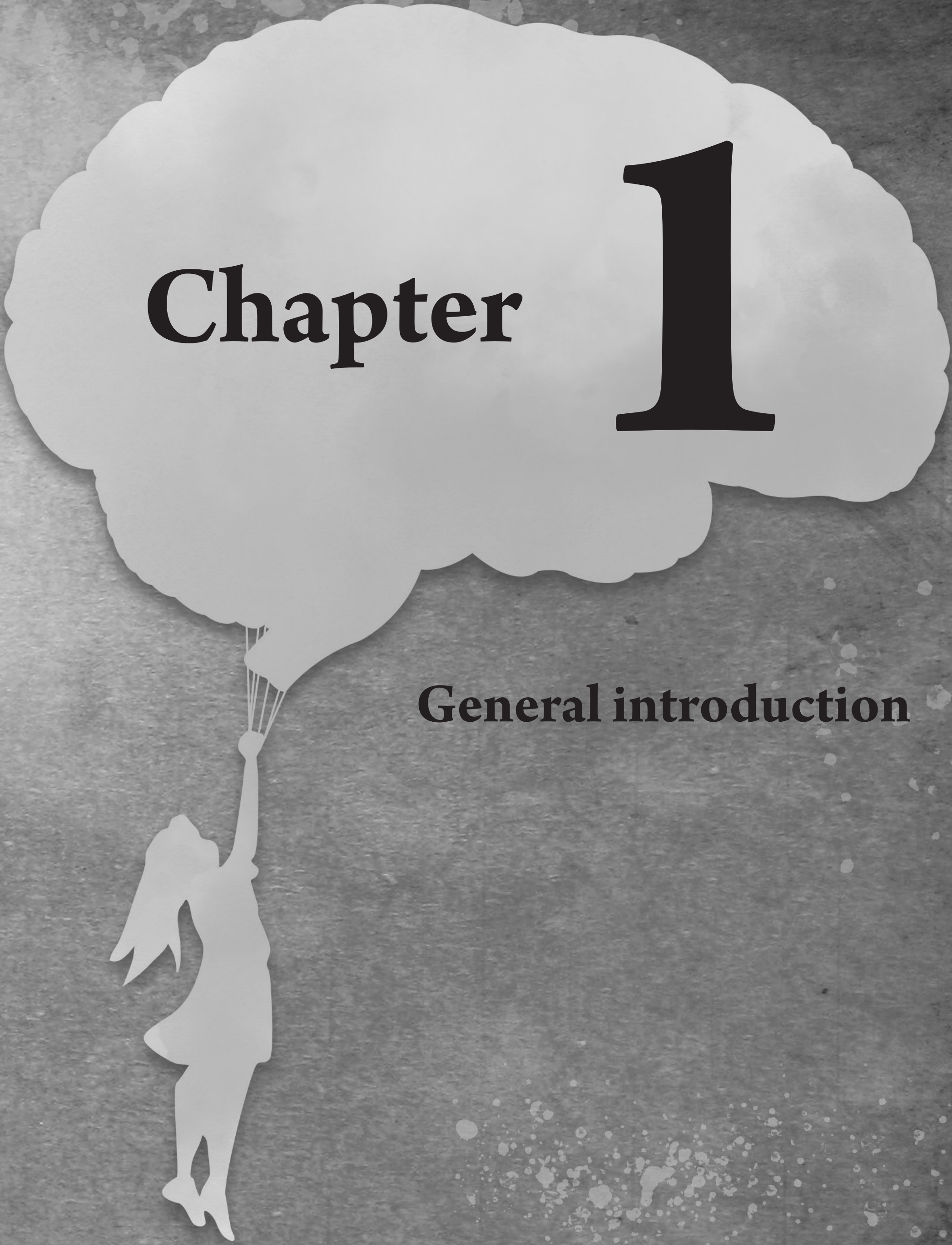
Copromotor:
Dr. R. Huffmeijer

Promotiecommissie:
Prof. dr. E.A.M. Crone
Prof. dr. C. Kemner (Universiteit Utrecht)
Dr. M.J. Peltola (University of Tampere)

This research was supported by the Leiden Consortium on Individual Development, which is funded through the Gravitation program of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant no. 024.001.003; MJB-K: VICI grant no. 453-09-003; MHvIJ: SPINOZA prize) and the European Research Council (MJB-K: ERC AdG 669249).

Contents

Chapter 1



General introduction

General introduction

Humans are social creatures, and as such, social interactions crucially affect our subjective well-being and happiness (Helliwell, 2003; Helliwell & Putnam, 2004; Pichler, 2006). Many factors, including verbal and non-verbal communication skills, emotion regulation abilities, and face processing capacities, may affect how effective we are in interacting with others. Face processing enables individuals to determine another person's gender, emotional state, and their degree of familiarity with another person. As a consequence, humans are able to adjust their behavior quickly based on who they are interacting with. However, individual differences may exist in how people process, and, ultimately, react to faces. Differences in temperament, the context in which a face is perceived, and (childhood) experiences with others may for instance influence how people react to faces. In the current thesis, the potential influence of some of these factors on the neural processing of and behavioral reactions to (child) faces, are discussed.

Processing of familiar and unfamiliar faces

Previous research has shown that newborns are already able to discriminate between their mother's and a stranger's face (Bushnell, Sai & Mullin, 1989; Bushnell, 2001; Field Cohen, Garcia & Greenberg, 1984; Walton, Bower & Bower, 1992), suggesting that basic face processing abilities are innate to some degree. Face processing capacities further develop through infancy, childhood and adolescence (see Pascalis et al., 2011) until face processing expertise is reached in adulthood, reflected by adults' capacity to remember hundreds of different faces over a long period of time, and their ability to distinguish between highly resembling faces (Bahrick, Bahrick & Wittlinger, 1975).

Face processing is a complex process that involves multiple brain areas, and several factors, such as individuals' degree of familiarity with another person, may also affect how people react to a particular face (see for a review Natu & O'Toole, 2011). The core face processing areas include the occipital face area, the fusiform face area (FFA), and the posterior superior temporal sulcus (STS). The occipital face area encodes information about the different parts of faces, such as eyes and mouth, and reacts to small changes in physical facial features (Liu, Harris, & Kanwisher, 2010; Nichols, Betts, & Wilson, 2010; Pitcher, Walsh, Yovel, & Duchaine 2007). Subsequently, more complex processing of facial features occurs in the lateral fusiform gyrus (FFA), associated with analyzing the invariant features of faces, such as gender and identity, and in the posterior STS, involved in the processing of the changeable features of a face, such as facial expressions and eye gaze (Andrews & Ewbank, 2004; Andrews & Schluppeck, 2004; Grill-Spector, Knouf & Kanwisher, 2004; Kanwisher & Yovel, 2006, Gobbini & Haxby, 2007; Yovel & Kanwisher, 2005).

In addition, brain areas involved in social and cognitive functions, such as Theory of Mind (e.g. anterior cingulate cortex), also show enhanced activity in response to faces, especially more familiar faces (Gobbini & Haxby, 2007). In reaction to familiar faces, enhanced activity is also frequently seen in brain areas involved in the retrieval of another person's biographical information, such as the anterior temporal cortex, brain areas related to the retrieval of memories of shared experiences (e.g. precuneus and cuneus), and brain areas associated with emotional responses to stimuli, including faces, such as amygdala and insula (Dubois et al., 1999; Gobbini, Leibenluft, Santiago, & Haxby 2004; Leibenluft, Gobbini, Harrison, & Haxby 2004; Schwartz et al., 2003; Sugiura et al., 2001). Thus, more familiar faces generally evoke enhanced activity in a more widespread network of brain areas than less familiar or unfamiliar faces, probably induced by the accumulation of experiences and social interactions people have with (highly) familiar individuals over time (Balas, Cox, & Conwell 2007).

Facial resemblance and kinship

Face processing also enables individuals to identify genetic relatives based on facial resemblance (Alvergne, Faurie, & Raymond, 2009; Bressan & Grassi, 2004; Kaminski, Dridi, Graff, & Gentaz, 2009; Lieberman, Tooby, & Cosmides 2007; Maloney & Dal Martello, 2006). Humans may favor individuals that facially resemble themselves over individuals that do not, because of the suspected genetic relatedness. In previous research, adult participants' were for instance more willing to cooperate with adults that facially resembled themselves (DeBruine et al., 2011; Krupp et al., 2008). In addition, facial resemblance increases 'parental' responses, like the willingness to invest in a child (Bressan, Bertamini, Nalli, & Zanutto, 2009; DeBruine, 2004; Platek et al., 2004). Editing children's pictures to make them facially resemble the participants may therefore offer an opportunity to simulate an 'own child' in individuals without children of their own. The suggested biological relatedness implied by the child's facial resemblance with the participant may trigger caregiving reactions in participants without children of their own.

A parent's caregiving system is particularly activated when the child is in (potential) danger or distress (George & Salomon, 2008). Similarly, a participant's caregiving systems may be activated when a child that facially resembles the participant is threatened or in danger. Protective behaviors and related changes in brain activity may even be evoked when individuals are not consciously aware of a threat (Bowlby 1988; Bakermans-Kranenburg & Van IJzendoorn, 2017). Thus, threatening stimuli may be processed preconsciously to some degree (Almeida, Pajtas, Mahon, Nakayama, & Caramazza, 2013; Morris, Öhman & Dolan, 1998; Whalen et al., 1998). In Chapter 2 of this thesis, we investigate whether subliminally presented threatening primes evoke the expected changes in amygdala activity (a brain region related to the processing of emotional, especially threatening information [LeDoux, 1998]). We simulate genetic relatedness by morphing a picture of a child's face unfamiliar to the participant with the participant's own face. In addition to effects of facial resemblance in brain areas involved in face processing and social cognition (see above), we would therefore expect stronger (neural) protective responses when threatening primes precede child faces that resemble the participant's face compared to child faces that do not resemble the participant's face (since self-resembling, but not non-resembling child faces are expected to evoke caregiving reactions). The studies presented in this thesis provide a 'proof of concept' with the aim to get first insights in the neural processes related to protective caregiving reactions, and are therefore based on a homogenous sample of young-adult female participants without children of their own.

Individual differences: empathy and love withdrawal

When a (self-resembling) child is threatened or in danger, individuals' (neural and behavioral) reactions may also be influenced by their levels of empathy. Empathy describes the ability to experience and understand the emotional states of others (Eres, Decety, Louis, & Molenberghs, 2015), and can be divided into a cognitive perspective-taking (i.e. understanding what the other feels), and an affective (i.e. feeling what the other feels) aspect. High empathy in children is generally related to a range of positive outcomes, such as higher levels of prosocial behavior, and lower levels of aggression (Findlay, Girardi & Coplan, 2006; Hastings, Zahn-Waxler, Robinson, Usher, & Bridges, 2000;

Miller, Eisenberg, Fabes, & Shell, 1996), and empathic concern (an aspect of affective empathy) in adults seems to drive (costly) altruistic behaviors (Batson, Ahmad, Lishner, & Tsang, 2002; Feldman-Hall, Dalgeish, Evans, & Mobbs, 2015). Thus, individuals that score high on empathic concern may respond stronger and may be more likely to engage in altruistic behavior when observing a child that is threatened or in danger.

Generally, individuals may be more likely to behave empathically and altruistically towards in-group members than towards out-group members (Cikara, Bruneau & Saxe, 2011; Levine, Prosser, Evans, & Reicher, 2005). In the context of the current thesis, the self-resembling child faces represent an in-group member (due to the implied close genetic relatedness), whereas the non-resembling child faces may be considered an out-group member (i.e. no [close] genetic relatedness with the participant implied). In previous research, participants' self-reported scores on empathic concern were related to anterior insula activity, which was more enhanced for in-group members than for out-group members. Anterior insula activity and associated scores on empathic concern also predicted how likely participants were to help other individuals in distress (Hein, Silani, Preuschoff, Batson, & Singer, 2010).

In addition to individuals' levels of empathy, reactions to self-resembling child faces, may also be affected by individuals' own childhood experiences with caregivers. Early experiences with caregivers may shape an individual's beliefs about relationships with and responses to significant others, such as close family members (Mikulincer, Shaver, Gillath, & Nitzberg, 2005), and may possibly also affect their reactions to self-resembling child faces, because of the suggested genetic relatedness. In general, relationships with early caregivers may have a profound impact on children's development. Insecurely attached children (generally associated with insensitive caregiving), score for instance lower on academic and social skills, and have on average more externalizing problems than securely attached children (Groh, et al, 2014; Groh, Fearon, Van IJzendoorn, Bakermans-Kranenburg, & Roisman, 2017; Kerns & Brumario, 2016; Williford, Carter & Pianta, 2016).

Negative parenting styles, such as psychological control (i.e. inducing guilt or shame, or making love conditional; Barber 1996) and harsh control (i.e. physical or verbal punishment) are also associated with more externalizing problems in children and adolescents (see Pinquart, 2017). Love withdrawal, an aspect of psychological control, in which the parent's love and affection become conditional on the child's behavior and success, is associated with enhanced anxiety, depressive symptoms, and lower self-control in children and adolescents (Hill & Bush, 2001; Mandara & Pikes, 2008). The effects of early experiences with caregivers may persist into adulthood. Experiences with psychological control during childhood are for instance related to insecure attachment and fear of failure in adults (Elliot & Thrash, 2004; Swanson & Malinckrodt, 2001). In addition, childhood experiences with parental love withdrawal may affect how young adults process and react to socio-emotional information, such as faces (Huffmeijer, Tops, Alink, Bakermans-Kranenburg, & Van IJzendoorn 2011, Huffmeijer et al., 2013). Thus, negative experiences with early caregivers, including the frequent use of love withdrawal as a disciplinary strategy, may affect individuals' reactions to socially relevant stimuli, such as self-resembling child faces.

Chapter 2 and Chapter 3 of the current thesis focus on the effects of facial resemblance on the neural processing and appraisal of child faces. Neural processing is measured using functional magnetic resonance imaging (fMRI; Chapter 2). Moderating effects of participants' experiences of love withdrawal, and their scores on empathic concern on the neural processing of child faces that differ in their degree of resemblance with the participant's faces are examined (Chapter 2). In addition, participants' appraisal of these child faces with differing degrees of self-resemblance is measured on a range of positive and negative criteria (Chapter 3). In Chapter 3, we also investigate whether love withdrawal moderates participants' evaluations of these child faces, and whether effects of facial resemblance depend on participants' neural processing of facial identity, as indicated by (the level of) FFA activity.

Test-retest reliability

Thorough conclusions can only be drawn from fMRI-research when the acquired data is valid and reliable. However, only a few studies have previously investigated the test-retest reliability of fMRI activity elicited with face processing tasks. In Chapter 4 of this thesis, we therefore examine the reliability of task-fMRI data acquired for the (non-emotional) face processing task used in the studies presented in this thesis, with the aim to investigate whether we can reliably measure changes in brain activity with our face processing paradigm. In case of low test-retest reliabilities, results obtained with our face processing task should be interpreted with caution, particular those relating to individual differences. Increasing the number of trials of task-fMRI may improve reliability (Bennett & Miller, 2010), and could thus ensure acceptable test-retest reliability of fMRI data. Therefore, we also examine the effect of increasing the number of trials of our research paradigm on test-retest-reliability estimates. In addition, we investigate the influence of participants' handedness on the reliability of participants' fMRI data. Left-handed individuals are frequently excluded from fMRI research, although they represent about ten percent of the human population (Willems, Van der Haegen, Fisher, & Francks, 2014; McManus, 2009). Therefore, including left-handed individuals in fMRI research appears desirable to us, and could be supported by acceptable reliability estimates of fMRI data acquired from left-handed participants.

Aims and outline of this thesis

The general aim of the current thesis is to increase our knowledge of individual differences in the neural processing and appraisal of children's faces that differ in their degree of resemblance with the participant's face. The central question we aim to answer is whether individual differences in early parenting experiences (i.e. love withdrawal), empathy and FFA activity on the one hand, and the context in which the child faces are presented on the other, may affect brain responses to and appraisal of self-resembling child faces. We investigate whether the degree of resemblance of a child's face with the participant's face affects how participants' process and evaluate these child faces. In addition, we examine whether participants' experiences of love withdrawal, current levels of empathy, FFA activity, and the context in which the child faces are presented, moderate the effects of facial resemblance. Figure 1 illustrates the topics that are discussed in the current thesis.

Chapter 2 focuses on participants' neural responses to child faces that differ in their degree of resemblance with the participant's face, both in neutral and threatening contexts. Moderating effects of love withdrawal and empathy are examined to explore associations between participants' experiences with love withdrawal, their levels of empathy and their neural processing of facial resemblance. In Chapter 3, we focus on participants' appraisal of these child faces. More specifically, we examine how participants' evaluations of the different child faces on a range of negative and positive criteria are affected by children's degree of resemblance with the participant's face. In addition, we explore how experiences of love withdrawal and the extent of neural face processing (i.e. FFA activity) moderate participants' appraisal of these child faces. Thus, we aim to increase our insight in correlations between brain (FFA activity) and behavior (appraisal of the different child faces). Chapter 4 focuses on the test-retest reliability of the fMRI data acquired during the face processing paradigm we used to examine the effects of facial resemblance on participants' brain activity. We estimate reliability of fMRI activity in several regions of interest for different numbers of trials, and we examine the influence of participants' handedness on the reliability of participants' fMRI data. In Chapter 5, we discuss our findings, elaborate on the limitations of the studies, and discuss implications for future research.

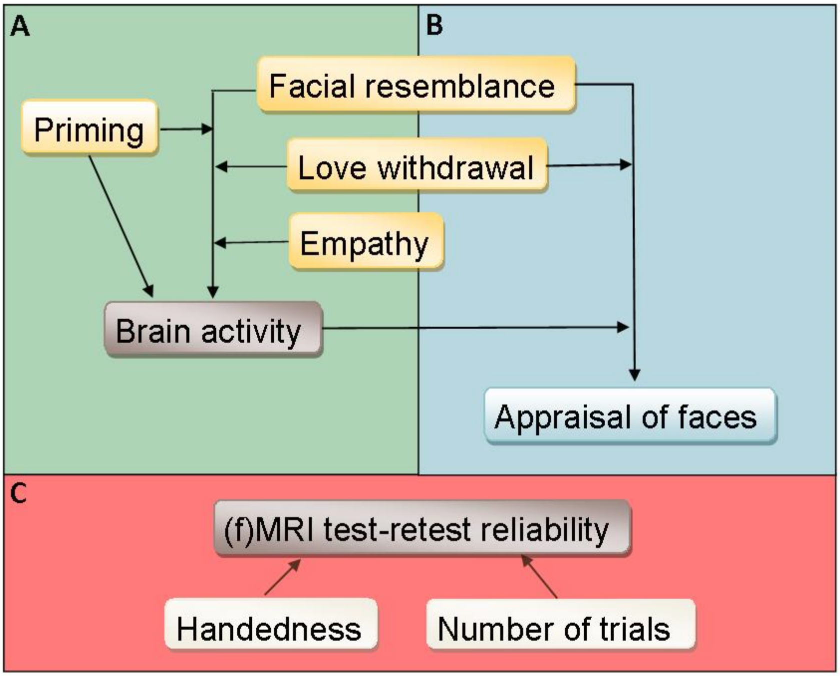


Figure 1. Graphic representation of the topics presented in the current thesis. In Chapter 2 (A), we examine the processing of subliminally presented threatening primes, their effects on neural responses to child faces that either resemble or do not resemble the participant's face, and moderating effects of empathy and experiences of love withdrawal. In Chapter 3 (B), we investigate the influence of the degree of resemblance of children's faces with the participant's face on participant's positive and negative appraisal of the child faces, and moderating effects of love withdrawal and participants' neural processing of facial identity (FFA activity). In Chapter 4 (C), we examine the test-retest reliability of the face processing task included in this thesis, and potential effects of handedness and the number of trials on reliability estimates.

References

Almeida, J., Pajtas, P. E., Mahon, B. Z., Nakayama, K., and Caramazza, A. (2013). Affect of the unconscious: Visually suppressed angry faces modulate our decisions. *Cogn. Affect. Behav. Neurosci.* 13, 94–101. doi: 10.3758/s13415-012-0133-7

Alvergne, A., Faurie, C., & Raymond, M. (2009). Father-offspring resemblance predicts paternal investment in humans. *Animal Behaviour, 78*(1), 61-69. doi: 10.1016/j.anbehav.2009.03.019

Andrews, T. J., & Ewbank, M. P. (2004). Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *Neuroimage, 23*(3), 905-913. doi:10.1016/j.neuroimage.2004.07.060

Andrews, T. J., & Schluppeck, D. (2004). Neural responses to Mooney images reveal a modular representation of faces in human visual cortex. *Neuroimage, 21*(1), 91-98. doi: 10.1016/j.neuroimage.2003.08.023

Bahrick, H. P., Bahrick, P. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of experimental psychology: General, 104*(1), 54. doi: 10.1037/0096-3445.104.1.54

Bakermans-Kranenburg, M. J. & Van IJzendoorn, M. H. (2017). Oxytocin and Human Sensitive and Protective Parenting. *Current topics in behavioral neurosciences.*

Balas, B., Cox, D., & Conwell, E. (2007). The effect of real-world personal familiarity on the speed of face information processing. *PloS one, 2*(11), e1223. doi: 10.1371/journal.pone.0001223

Barber, B. K. (1996). Parental psychological control: Revisiting a neglected construct. *Child development, 67*(6), 3296-3319. doi: 10.1111/j.1467-8624.1996.tb01915.x

Batson, C. D., Ahmad, N., Lishner, D. A., & Tsang, J. (2002). Empathy and altruism. In *The Oxford handbook of hypo-egoic phenomena*.

Bennett, C. M., & Miller, M. B. (2010). How reliable are the results from functional magnetic resonance imaging?. *Annals of the New York Academy of Sciences, 1191*(1), 133-155. doi: 10.1111/j.1749-6632.2010.05446.x

Bowlby, J. (1988). "Attachment, communication and the therapeutic process", in *A Secure Base Parent-Child Attachment and Healthy Human Development,* (London: Routledge), 137-157.

Bressan, P., and Grassi, M. (2004). Parental resemblance in 1-years-olds and the Gaussian curve. *Evol. Hum. Behav.* 25, 133–141. doi: 10.1016/j.evolhumbehav.2004.03.001

Bressan, P., Bertamini, M., Nalli, A., & Zanutto, A. (2009). Men do not have a stronger preference than women for self-resemblant child faces. *Archives of sexual behavior, 38*(5), 657-664. doi: 10.1007/s10508-008-9350-0

Bushnell, I. W. R., Sai, F., & Mullin, J. T. (1989). Neonatal recognition of the mother's face. *British Journal of Developmental Psychology, 7*(1), 3-15. doi: 10.1111/j.2044-835X.1989.tb00784.x

Bushnell, I. W. R. (2001). Mother's face recognition in newborn infants: Learning and memory. *Infant and Child Development, 10*(1-2), 67-74. doi: 10.1002/icd.248

Cikara, M., Bruneau, E. G., & Saxe, R. R. (2011). Us and them intergroup failures of empathy. *Current Directions in Psychological Science, 20*(3), 149-153. doi: 10.1177/0963721411408713

DeBruine, L. M. (2004). Resemblance to self increases the appeal of child faces to both men and women. *Evolution and Human Behavior, 25*(3), 142-154. doi: 10.1016/j.evolhumbehav.2004.03.003

DeBruine, L. M., Jones, B. C., Watkins, C. D., Roberts, S. C., Little, A. C., Smith, F. G., & Quist, M. C. (2011). Opposite-sex siblings decrease attraction, but not prosocial attributions, to self-resembling opposite-sex faces. *Proceedings of the National Academy of Sciences, 108*(28), 11710-11714. doi: 10.1073/pnas.1105919108

Dubois, S., Rossion, B., Schiltz, C., Bodart, J. M., Michel, C., Bruyer, R., & Crommelinck, M. (1999). Effect of familiarity on the processing of human faces. *Neuroimage, 9*(3), 278-289. doi: 10.1006/nimg.1998.0409

Elliot, A. J., & Thrash, T. M. (2004). The intergenerational transmission of fear of failure. *Personality and Social Psychology Bullitin.* 30(8), 957–971. doi: 10.1177/0146167203262024

Eres, R., Decety, J., Louis, W. R., and Molenberghs, P. (2015). Individual differences in local gray matter density are associated with differences in affective and cognitive empathy. *Neuroimage* 117, 305–310. doi: 10.1016/j.neuroimage.2015.05.038

FeldmanHall, O., Dalgleish, T., Evans, D., & Mobbs, D. (2015). Empathic concern drives costly altruism. *Neuroimage, 105*, 347-356. doi: 10.1016/j.neuroimage.2014.10.043

Field, T. M., Cohen, D., Garcia, R., & Greenberg, R. (1984). Mother-stranger face discrimination by the newborn. *Infant Behavior and development, 7*(1), 19-25. doi: 10.1016/S0163-6393(84)80019-3

Findlay, L. C., Girardi, A., & Coplan, R. J. (2006). Links between empathy, social behavior, and social understanding in early childhood. *Early Childhood Research Quarterly, 21*(3), 347-359. doi: 10.1016/j.ecresq.2006.07.009

George, C., and Solomon, J. (2008). "The caregiving system: a behavioral systems approach to parenting", in *Handbook of Attachment: Theory Research and Clinical Applications*, 2nd Edn, eds J. Cassidy and P. R. Shaver (New York, NY: Guilford Press), 833–856.

Gobbini, M. I., & Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia, 45*(1), 32-41. doi : 10.1016/j.neuropsychologia.2006.04.015

Gobbini, M. I., Leibenluft, E., Santiago, N., & Haxby, J. V. (2004). Social and emotional attachment in the neural representation of faces. *Neuroimage, 22*(4), 1628-1635. doi : 10.1016/j.neuroimage.2004.03.049

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature neuroscience, 7*(5), 555.doi: 10.1038/nn1224

Groh, A. M., Fearon, R. P., Bakermans-Kranenburg, M. J., Van IJzendoorn, M. H., Steele, R. D., & Roisman, G. I. (2014). The significance of attachment security for children's social competence with peers: A metaanalytic study. *Attachment and Human Development, 16,* 103–136. doi: 10.1080/14616734.2014.883636

Groh, A. M., Fearon, R. M., IJzendoorn, M. H., Bakermans-Kranenburg, M. J., & Roisman, G. I. (2017). Attachment in the early life course: Meta-analytic evidence for its role in socioemotional development. *Child Development Perspectives, 11*(1), 70-76. doi: 10.1111/cdep.12213

Hastings, P. D., Zahn-Waxler, C., Robinson, J., Usher, B., & Bridges, D. (2000). The development of concern for others in children with behavior problems. *Developmental psychology, 36*(5), 531. doi: 10.1037/0012-1649.36.5.531

Hein, G., Silani, G., Preuschoff, K., Batson, C. D., & Singer, T. (2010). Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. *Neuron, 68*(1), 149-160. doi: 10.1016/j.neuron.2010.09.003

Helliwell, J. F. (2003). How's life? Combining individual and national variables to explain subjective well-being. *Economic Modelling, 20,* 331–360. doi: 10.1016/S0264-9993(02)00057-3

Helliwell, J. F., & Putnam, R. (2004). The social context of well-being. *Philosophical Transactions of the Royal Society London, 359,* 1435–1446. doi: 10.1098/rstb.2004.1522

Hill, N. E., & Bush, K. R. (2001). Relationships between parenting environment and children's mental health among African American and European American mothers and children. *Journal of Marriage and Family, 63*(4), 954-966. doi: 10.1111/j.1741-3737.2001.00954.x

Hoffmann, M., Lipka, J., Mothes-Lasch, M., Miltner, W. R., and Straube, T. (2012). Awareness modulates response of the amygdala and the visual cortex to highly arousing visual threat. *Neuroimage* 62, 1439–1444. doi: 10.1016/j.neuroimage.2012.05.064

Huffmeijer, R., Tops, M., Alink, L. R., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2011). Love withdrawal is related to heightened processing of faces with emotional expressions and incongruent emotional feedback: evidence from ERPs. *Biological Psychology, 86*(3), 307-313. doi: 10.1016/j.biopsycho.2011.01.003

Huffmeijer, R., Alink, L. R., Tops, M., Grewen, K. M., Light, K. C., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2013). The impact of oxytocin administration and maternal love withdrawal on event-related potential (ERP) responses to emotional faces with performance feedback. *Hormones and behavior, 63*(3), 399-410. doi: 10.1016/j.yhbeh.2012.11.008

Kaminski, G., Dridi, S., Graff, C., & Gentaz, E. (2009). Human ability to detect kinship in strangers' faces: effects of the degree of relatedness. *Proceedings of the Royal Society of London B: Biological Sciences, 276*(1670), 3193-3200. doi: 10.1098/rspb.2009.0677

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 361*(1476), 2109-2128. doi: 10.1098/rstb.2006.1934

Kerns, K. A. & Brumariu, L. E. (2016). Attachment in middle childhood. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (pp. 349-365). New York, NY: The Guilford Press.

Krupp, D. B., DeBruine, L. M., & Barclay, P. (2008). A cue of kinship promotes cooperation for the public good. *Evolution and Human Behavior, 29*(1), 49-55. doi: 10.1016/j.evolhumbehav.2007.08.002

LeDoux, J. (1998). Fear and the brain: where have we been, and where are we going? *Biol. Psychiatry* 44, 1229–1238. doi: 10.1016/s0006-3223(98)00282-0. https://doi.org/110.1016/S0006-3223(98)00282-0

Leibenluft, E., Gobbini, M. I., Harrison, T., and Haxby, J. V. (2004). Mothers' neural activation in response to pictures of their children and other children. *Biol. Psychiatry* 56, 225–232. doi: 10.1016/j.biopsych.2004.05.017

Levine, M., Prosser, A., Evans, D., & Reicher, S. (2005). Identity and emergency intervention: How social group membership and inclusiveness of group boundaries shape helping behavior. *Personality and Social Psychology Bulletin, 31*(4), 443-453. doi: 10.1177/0146167204271651

Lieberman, D., Tooby, J., & Cosmides, L. (2007). The architecture of human kin detection. *Nature, 445*, 727–731. doi: 10.1038/nature05510

Liu, J., Harris, A., & Kanwisher, N. (2010). Perception of face parts and face configurations: an fMRI study. *Journal of cognitive neuroscience, 22*(1), 203-211. doi: 10.1162/jocn.2009.21203

Maloney, L. T., & Dal Martello, M. F. (2006). Kin recognition and the perceived facial similarity of children. *Journal of Vision, 6*(10), 4-4. doi: 10.1167/6.10.4

Mandara, J., & Pikes, C. L. (2008). Guilt trips and love withdrawal: Does mothers' use of psychological control predict depressive symptoms among African American adolescents?. *Family Relations, 57*(5), 602-612. doi: 10.1111/j.1741-3729.2008.00526.x

McManus, I. C. (2009). The history and geography of human handedness. *Language lateralization and psychosis*, 37-57. doi: 10.1038/nrn3679

Mikulincer, M., Shaver, P. R., Gillath, O., and Nitzberg, R. A. (2005). Attachment, caregiving and altruism: boosting attachment security increases compassion and helping. *J. Pers. Soc. Psychol.* 89, 817–839. doi: 10.1037/0022-3514.89.5.817

Miller, P. A., Eisenberg, N., Fabes, R. A., & Shell, R. (1996). Relations of moral reasoning and vicarious emotion to young children's prosocial behavior toward peers and adults. *Developmental Psychology, 32*(2), 210. doi: 10.1037/0012-1649.32.2.210

Morris, J. S., Öhman, A., and Dolan, R. J. (1998). Conscious and unconscious emotional learning in the human amygdala. *Nature* 393, 467–470. doi: 10.1038/30976

Natu, V., & O'Toole, A. J. (2011). The neural processing of familiar and unfamiliar faces: a review and synopsis. *British Journal of Psychology, 102*(4), 726-747. doi: 10.1111/j.2044-8295.2011.02053.x

Nelson, C. A., & Ludemann, P. M. (1989). Past, current, and future trends in infant face perception research. *Canadian Journal of Psychology/Revue canadienne de psychologie, 43*(2), 183. doi: 10.1037/h0084221

Nichols, D. F., Betts, L. R., & Wilson, H. R. (2010). Decoding of faces and face components in face-sensitive human visual cortex. *Frontiers in Psychology, 1*. doi: 10.3389/fpsyg .2010.00028

Pascalis, O., de Martin de Viviés, X., Anzures, G., Quinn, P. C., Slater, A. M., Tanaka, J. W., & Lee, K. (2011). Development of face processing. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*(6), 666-675. doi: 10.1002/wcs.146

Pessoa, L., Japee, S., Sturman, D., and Ungerleider, L. G. (2006). Target visibility and visual awareness modulate amygdala responses to fearful faces. *Cereb. Cortex* 16, 366–375. doi: 10.1093/cercor/bhi115

Pichler, F. (2006). Subjective quality of life of young Europeans. Feeling happy but who knows why? *Social Indicators Research, 75,* 419–444. doi: 10.1007/s11205-004-4337-2

Pinquart, M. (2017). Associations of parenting dimensions and styles with externalizing problems of children and adolescents: An updated meta-analysis. *Developmental psychology,* 53(5), 873. doi: 10.1037/dev0000295

Pitcher, D., Walsh, V., Yovel, G., & Duchaine, B. (2007). TMS evidence for the involvement of the right occipital face area in early face processing. *Current Biology, 17*(18), 1568-1573. doi: 10.1016/j.cub.2007.07.063

Platek, S. M., Raines, D. M., Gallup, G. G., Mohamed, F. B., Thomson, J. W., Myers, T. E., ... & Arigo, D. R. (2004). Reactions to children's faces: Males are more affected by resemblance than females are, and so are their brains. *Evolution and Human behavior, 25*(6), 394-405. doi: 10.1016/j.evolhumbehav.2004.08.007

Schwartz, C. E., Wright, C. I., Shin, L. M., Kagan, J., Whalen, P. J., McMullin, K. G., & Rauch, S. L. (2003). Differential amygdala response to novel versus newly familiar neutral faces: A functional MRI probe developed for studying inhibited temperament. *Biological Psychiatry, 53,* 854–862. doi: 10.1016/S0006-3223(02)01906-6

Sugiura, M., Kawashima, R., Nakamura, K., Sato, N., Nakamura, A., Kato, T., ... & to, K. (2001). Activation reduction in anterior temporal cortices during repeated recognition of faces of personal acquaintances. *Neuroimage, 13*(5), 877-890. doi: 10.1006/nimg.2001.0747

Swanson, B., & Mallinckrodt, B. (2001). Family environment, love withdrawal, childhood sexual abuse, and adult attachment. *Psychotherapy Research, 11*(4), 455-472.

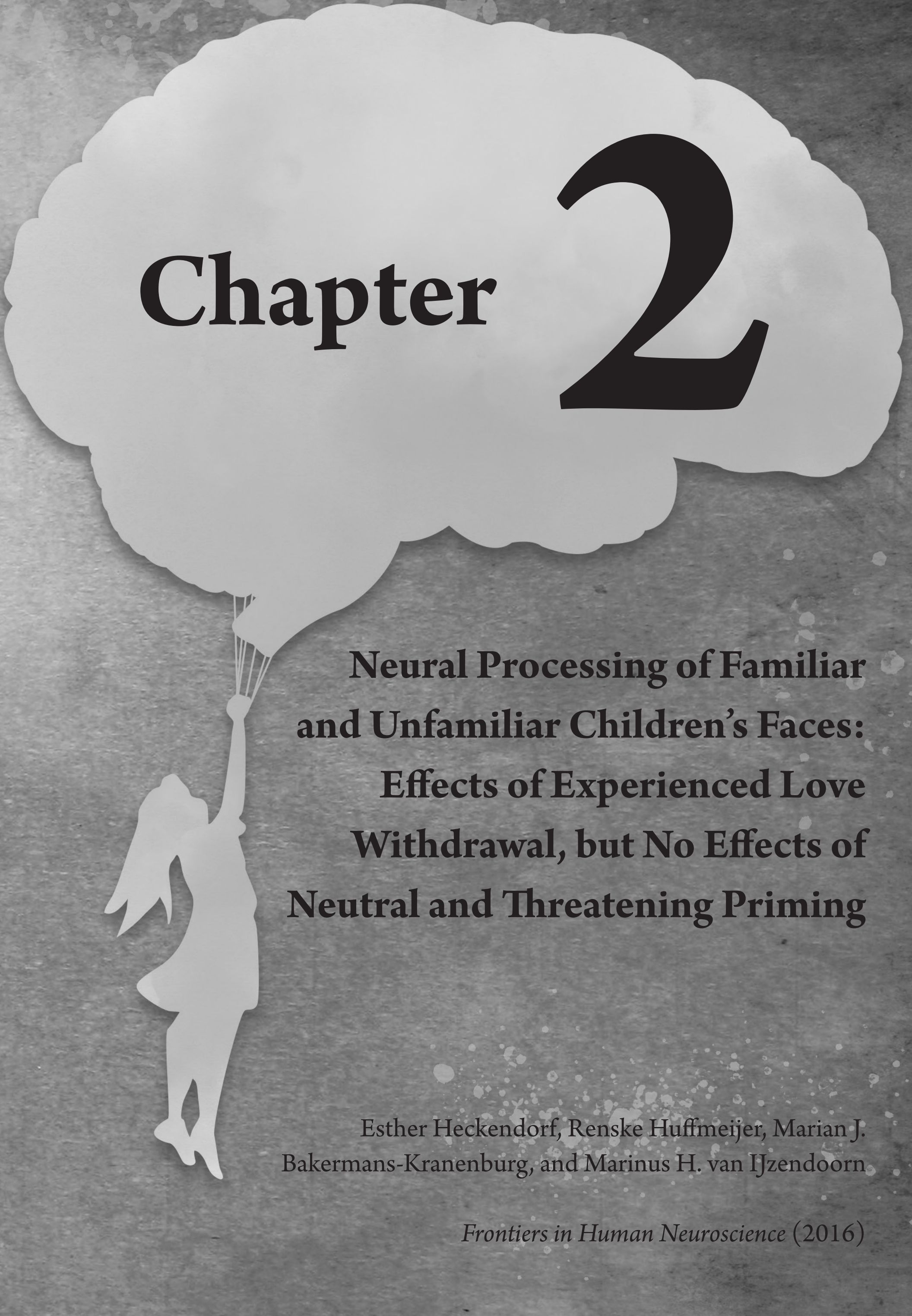
Walton, G. E., Bower, N. J. A., & Bower, T. G. R. (1992). Recognition of familiar faces by newborns. *Infant Behavior and Development, 15*(2), 265-269. doi: 10.1016/0163-6383(92)80027-R

Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., and Jenike, M. A. (1998). Masked presentation of emotional facial expressions modulate amygdala activity without explicit knowledge. *J. Neurosci.* 18, 411–418.

Willems, R. M., Van der Haegen, L., Fisher, S. E., & Francks, C. (2014). On the other hand: including left-handers in cognitive neuroscience and neurogenetics. Nature Reviews. *Neuroscience, 15*(3), 193. doi: 10.1038/nrn3679

Williford, A. P., Carter, L. M., & Pianta, R. C. (2016). Attachment and school readiness. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (pp. 966-982). New York, NY: The Guilford Press.

Yovel, G., & Kanwisher, N. (2005). The neural basis of the behavioral face-inversion effect. *Current Biology, 15*(24), 2256-2262. doi: 10.1016/j.cub.2005.10.072



Chapter 2

Neural Processing of Familiar and Unfamiliar Children's Faces: Effects of Experienced Love Withdrawal, but No Effects of Neutral and Threatening Priming

Esther Heckendorf, Renske Huffmeijer, Marian J. Bakermans-Kranenburg, and Marinus H. van IJzendoorn

Frontiers in Human Neuroscience (2016)

Abstract

In the face of a potential threat to his or her child, a parent's caregiving system becomes activated, motivating the parent to protect and care for the child. However, the neural correlates of these responses are not yet well understood. The current study was a pilot study to investigate the processing of subliminally presented threatening primes and their effects on neural responses to familiar and unfamiliar children's faces. In addition, we studied potential moderating effects of empathy and childhood experiences of love-withdrawal. A total of 45 students participated in an fMRI experiment in which they were shown pictures of familiar children (pictures morphed to resemble the participant like an own child would) and unfamiliar children preceded by neutral and threatening primes. Participants completed a modified version of the Children's Report of Parental Behavior Inventory to measure parental love withdrawal, and the Empathic Concern scale of the Interpersonal Reactivity Index to measure affective empathy. Contrary to our expectations, we did not find evidence for subliminal priming effects. However, we did find enhanced activity in the right inferior frontal gyrus (IFG; involved in self-referential processing) and in face processing areas (infero-lateral occipital cortex and fusiform areas) in response to the familiar child, indicating preferential processing of these faces. Effects of familiarity in face processing areas were larger for participants reporting more love withdrawal, suggesting enhanced attention to and processing of these highly attachment relevant stimuli. Unfamiliar faces elicited enhanced activity in bilateral superior temporal gyrus (STG) and other regions associated with theory of mind (ToM), which may indicate more effortful ToM processing of these faces. We discuss the potential difference between a familiarity and a caregiving effect triggered by the morphed faces, and emphasize the need for replication in parents with pictures of their "real" own child.

Introduction

In the face of a potential threat or danger in the environment, a parent's caregiving system may become activated when his or her child or a stimulus reminiscent of that child (such as crying or a picture of the child's face) is present and the threat is not overwhelmingly strong (Mikulincer, Shaver, Gillath, & Nitzberg, 2005; George & Solomon, 2008; Swain et al., 2014). Even when a parent is not consciously aware of a threatening stimulus in the environment, he or she might still process this threatening stimulus to some extent, which could lead to specific parental behaviors (with accompanying changes in brain activity) to protect and care for the child (Bowlby, 1988; Bakermans-Kranenburg & Van IJzendoorn, 2017). It has been argued that the caregiving system is complementary to the attachment system (George & Solomon, 2008; Strathearn, Fonagy, Amico, & Montague, 2009), and is not restricted to the parent-child relationship but rather extends to other intimate relationships such as the relationships with siblings or partners (e.g., Mikulincer et al., 2005). In the current study we focus on the neural processing of familiar and unfamiliar faces after subliminal neutral or threatening primes. The familiar faces were created by morphing a child's face with the participant's own face to suggest familiarity and potentially biological relatedness in order to trigger the caregiving system.

Individuals may be able to process affective information, especially potentially threatening stimuli, fast and automatically, and possibly even without conscious awareness (Whalen et al., 1998; Globisch, Hamm, Esteves, & Ohman, 1999; Mikulincer et al., 2005). Since it may take hundreds of milliseconds to consciously perceive a potential threat (Koch & Tsuchiya, 2007), a system in the human brain that can react to potential threats before conscious awareness seems advantageous from an evolutionary perspective, as it enables a fast reaction that can preserve oneself or one's offspring from danger or death. Subliminal primes can be used to examine the preconscious processing of threat-related information. In some previous studies, researchers found evidence for the human brain's capacity to process threat-related visual stimuli without conscious awareness. For example, in one study participants rated neutral stimuli (the target) more positively when these stimuli were preceded by a subliminal prime depicting a happy face and more negatively when targets were preceded by a prime depicting an angry face (Almeida, Pajtas, Mahon, Nakayama, & Caramazza, 2013). Brain imaging studies also found some evidence for the brain's ability to process threatening stimuli without conscious awareness. In these studies, researchers mainly focused on amygdala activity in response to subliminally presented angry or fearful faces. The amygdala is a subcortical structure commonly associated with the processing of emotional, especially threat-related, content (LeDoux, 1998). Briefly presented fearful (Whalen et al., 1998) and angry (Morris, Öhman, & Dolan, 1998) faces evoked right amygdala activity.

However, in some studies no evidence for the existence of such an automatic processing system of threat-related stimuli was found. For example, in earlier studies with threat-related stimuli presented in supraliminal and subliminal conditions, enhanced amygdala activity was found in the supraliminal, but not in the subliminal condition (Pessoa, Japee, Sturman, & Ungerleider, 2006; Hoffmann, Lipka, Mothes-Lasch, Miltner, & Straube 2012). Importantly, not everyone may respond to emotional or threatening information in the same way, and such moderating effects may explain inconsistent findings for main effects of threat-related stimuli. Considering parental responses or responses to biologically related or otherwise familiar others in threatening contexts, factors such as empathy and individuals' own childhood experiences with their attachment figures may influence how they react to a potential threat to offspring or other familiar persons.

With regard to empathy, which has been defined as the capacity to experience and understand the emotional states of others (Eres et al., 2015), cognitive (understand), affective (experience) and imitative (action) components can be distinguished (Klimecki & Singer, 2013). In the current study, we are mainly interested in the affective component of empathy, which refers to how we feel when we imagine the emotions of another person in a particular situation (i.e., when we "put ourselves in the other person's shoes"). This affective component refers to a mature affective response that is experienced with a certain distance to the person empathized with rather than the more primitive and potentially dysfunctional copying of the target's affective response or distress (Davis, 1983; De Corte et al., 2007). In previous research, viewing a beloved person in pain elicited activity in brain areas associated with affective dimensions of pain (e.g., dorsal anterior cingulate cortex, dACC, see Lieberman & Eisenberger, 2015), with stronger effects in participants with high scores on empathic concern (Singer et al., 2004). In addition, observing someone experiencing "social pain" (i.e., being socially excluded) elicited brain activity in similar areas (e.g., anterior insula, anterior cingulate cortex) in highly empathic but not in less empathic participants (Masten, Morelli, & Eisenberger, 2011). Because pain, whether social or physical, results from a harmful stimulus in the environment, we may, extrapolating from these results, expect that highly empathic individuals will react stronger to a potential threat to their child or a familiar other. It should be noted, however, that the intensity of the threat could modulate responses of caregiving and protection, since overwhelmingly strong threats might turn the focus away from the other – even when it is offspring – to protecting oneself (Mikulincer et al., 2005). However, the stimuli used in the current study depict moderate rather than extreme threats.

Childhood experiences with parental love-withdrawal may also shape caregiving and protective responses to offspring or familiar others when confronted with a threat. Although the neural correlates of individual differences in caregiving and protective responses are poorly understood (but see Swain et al., 2014), the presence of a threat may affect the way parents perceive and respond to their child differently based on their own childhood experiences with protective or neglectful attachment figures. Love withdrawal is a parental disciplinary strategy in which the parent’s love and affection is conditional on the child’s behavior and success. Excessive use of love withdrawal is considered psychological maltreatment (Euser, Van IJzendoorn, Prinzie, & Bakermans-Kranenburg, 2010) and experiences of love withdrawal have been associated with long-lasting negative outcomes, like fear of failure, low self-esteem, low emotional well-being, and a negative view of parent-child relationships as well as insecure attachment (Bowlby, 1973/1985, p. 243; Assor, Roth, & Deci, 2004; Goldstein & Heaven, 2000; Elliot & Thrash, 2004; Renk, McKiney, Klein, & Oliveiros, 2006). Thus, experiencing love-withdrawal has consequences extending beyond the parent-child relationship, affecting ones beliefs about relationships as well as more generalized socio-emotional processes. That personal characteristics and belief systems formed within the parent-child relationship can affect responses to other significant others has convincingly been shown by, e.g., Mikulincer et al. (2005). These authors showed experimentally how feelings of more secure attachment facilitate supporting partners in distress. Previous research has associated childhood experiences of love withdrawal not only with changes in the (neural) processing of and responding to socio-emotional information, including faces (Huffmeijer et al., 2011), but also with changes in effects of external influences, including oxytocin administration, on these processes (Van IJzendoorn, Huffmeijer, Alink, Bakermans-Kranenburg, & Tops, 2011; Bakermans-Kranenburg, Van IJzendoorn, Riem, Tops, & Alink, 2012; Huffmeijer et al., 2013).

The present study was a pilot for research to be conducted with mothers, and examined in young-adult females without children of their own whether subliminally presented threatening primes would evoke the expected changes in brain activity in the amygdala and would differentially affect (the neural correlates of) protective responses to pictures of a familiar and an unfamiliar child. In addition, we examined whether these effects would be moderated by empathic concern and self-reported childhood experiences of love-withdrawal. In order to provide a "proof of concept", we used a homogenous student sample without children. We mimicked maternal reactions by presenting as "own child" the picture of a child face modified to resemble the participant's face, and combined this with primes depicting neutral and threatening scenes to evoke (the neural correlates of) protective responses. Facial resemblance is a very important cue for kinship (Bressan & Grassi, 2004; Maloney & Dal Martello, 2006) and has been shown to increase "parental" responses such as willingness to invest in a child (e.g., DeBruine, 2004; Platek et al., 2004). Thus, using pictures of children facially resembling the participants (by use of morphing, see "Materials and Methods" Section) is probably the most accurate imitation of an "own" child in participants without children of their own. However, we cannot exclude the possibility that the morphed faces will only be perceived as familiar rather than suggesting biological relatedness.

We focused our analyses on brain regions known to be involved in the processing of threat and face familiarity: the amygdala (involved in threat detection as well as more general salience detection, and responsive to face familiarity in previous studies [Natu & O'Toole, 2011]), inferior frontal gyrus (IFG, implicated in the processing of familiar faces, see for a review Devue & Brédart, 2011; Platek, Wathne, Tierny, & Thomson, 2008; implicated in affective empathy, Shamay-Tsoory, 2011, and considered part of the mirror neuron system, e.g., Kilner, Neal, Weiskopf, Friston, & Frith, 2009), and superior temporal gyrus (STG, found to be activated in response to unfamiliar compared to personally familiar faces, see Ramon, Vizioli, Liu-Shang, & Rossien 2015, and involved in Theory of Mind [ToM]). Importantly, these areas have not only been associated with the neural processing of threat and/or familiarity, but the functions mediated by these regions (such as ToM, empathy, affect regulation and mirroring) are also considered critical for parental behavior and involvement (Swain et al., 2014). We expected enhanced amygdala activity in response to threatening primes relative to neutral primes. We expected empathy to moderate this effect, with enhanced amygdala activity in highly empathic individuals. In addition, we hypothesized that IFG activity would be elevated in response to familiar-looking compared to unfamiliar-looking faces, and, conversely, that STG activity would be elevated in reaction to unfamiliar compared to familiar-looking faces. We explored potential moderating effects of experiences of love withdrawal, which might moderate effects of face familiarity or might be associated with the strength of a priming effect on familiar faces in particular. We chose to focus on a limited number of regions of interest (ROIs) to retain sufficient statistical power for testing *a priori* hypotheses, but, as interesting or unexpected effects might occur in other brain regions, we also conducted whole-brain analyses to explore changes in brain activity as a result of the primes, familiarity, empathy, and parental love withdrawal.

Materials and Methods

Participants

A total of 49 female undergraduate and graduate students aged 18–28 years (*M* = 21.73, *SD* = 2.55) were invited for two experimental sessions, separated by approximately 4 weeks. The second session was included to study test-retest reliability of fMRI data (to be reported elsewhere); the current study uses data from the first session only. Exclusion criteria were MRI contraindications, pregnancy, current psychiatric and neurological disorders, severe head injury, current alcohol or drug abuse, and chronic use of medication (except contraceptives). Data of four participants were excluded from analysis because of excessive head movements (>3 mm; *n* = 3) or falling asleep during fMRI acquisition (*n* = 1). Our final sample therefore included 45 participants with an average age of 21.82 years (*SD* = 2.61, range: 18–28). The study was approved by the Ethics Committee of the Leiden University Medical Center. All participants signed informed consent at the beginning of the first session and were rewarded with 40€ for participation. None of these participants' structural MRI scans showed any anomalies.

Procedure

Participants' handedness was assessed using van Strien's (1992) Handedness Questionnaire prior to the first session. Participants were asked to abstain from alcohol and excessive physical activity during the last 24 h and from caffeine during the last 12 h before the start of the session. At the beginning of the session participants completed questionnaires on empathy and parental use of love withdrawal. Subsequently, the MRI procedure was explained and participants were placed in the MRI scanner. Foam inserts were placed between the head coil and the participant's head to minimize head movements. Within the scanner, participants completed a priming task (see below), during which visual stimuli were projected onto a screen placed outside the opening of the scanner bore. Participants viewed the screen through a mirror fixed to the head coil. At the end of the second session participants were debriefed about the nature of the priming task.

Questionnaires

Handedness Questionnaire. This questionnaire consists of 10 items with regard to hand preference during execution of several tasks (e.g., "Which hand do you use to hold scissors?") scored on a 3-point scale (left hand, both hands, right hand) ranging from −1 to 1. Total scores can thus vary between −10 and +10. Individuals with a score of +8 or higher are classified as strongly right-handed, whereas individuals scoring −8 or lower are classified as strongly left-handed. Individuals with scores between −8 and +8 are classified as ambidexter (van Strien, 2003). According to this definition, in the current sample, 23 participants were strongly right-handed, 19 were strongly left-handed, and three were ambidexter. We oversampled left-handed participants in order to examine the potential influence of left-handedness on neural activity (to be reported elsewhere).

Children's Report of Parental Behavior Inventory. Participants completed a modified version of the 30-item Children's Report of Parental Behavior Inventory (CRPBI-30, Schludermann & Schludermann, 1983; Beyers & Goossens, 2003), containing the items of the Acceptance and Psychological Control scales from the original questionnaire and several extra items to measure love withdrawal. The 11-item Love Withdrawal scale consisted of all five items that constitute the Withdrawal of Relations subscale of the 108-item

CRPBI (3 of which are also included in the Psychological Control scale of the CRPBI-30; Schludermann & Schludermann, 1983), two items that were adapted from this same questionnaire, and four items that were adapted from the Parental Discipline Questionnaire (PDQ, Hoffman & Saltzstein, 1967; Patrick & Gibbs, 2007). Participants rated how well each item described their mother and father separately (e.g., "My mother was a person who if I'd hurt her feelings, stopped talking to me until I please her again") on a 5-point Likert scale, ranging from ("not at all") to ("very well"). We only included the 11-items of the Love Withdrawal subscale in our analyses. Scores for maternal and paternal love withdrawal were summed. After winsorizing the score of one outlier (z = 3.61; the new score was computed as the highest score occurring in the rest of the sample plus the difference between the highest and next-highest score, see Tabachnick and Fidell, 2001), the scores were normally distributed with an average score of 18.72 (SD = 6.15). Internal consistency of this questionnaire was high (Cronbach's alpha = 0.91). Adequate validity and reliability of the CRPBI and its subscales were demonstrated (Schludermann & Schludermann, 1983, 1988; Locke & Prinz, 2002) and the Love Withdrawal subscale as used in this study was implemented in earlier research on the consequences of maternal love withdrawal in young adults (Huffmeijer et al., 2011).

Interpersonal Reactivity Index. To measure empathy, participants completed the 28-item Interpersonal Reactivity Index, a well validated questionnaire measuring four distinct aspects of empathy (Perspective Taking, Fantasy, Empathic Concern, and Personal Distress; Davis, 1983; De Corte et al., 2007). In the current analyses, we only administered the seven-items of the Empathic Concern subscale, since we were interested in the emotional component of empathy. Participants rated how well each of the items described themselves on a 5-point Likert scale, ranging from 0 ("does not describe me well") to 4 ("describes me very well"). The data were normally distributed and did not contain any outliers. On average, participants scored 19.36 (SD = 3.53) on the Empathic Concern scale. The internal consistency was acceptable ($\alpha = 0.67$).

Scores on Love Withdrawal and Empathic Concern were not correlated (r = 0.00) and could therefore be included as independent predictors in the same analyses.

Experimental Task

In the scanner, subjects completed a priming task consisting of 234 trials. The priming task was set up in an event-related design. E-prime Software (Psychology Software Tools, 2012) was used for stimulus presentation. All stimuli were shown in the center of the screen on a black background. Forward and backward masking of the primes, using a picture showing a colored, circular pattern, was used on all trials to prevent conscious perception of the primes. The mask matched the dimensions and average luminosity of the primes. During each trial, a fixation cross was presented for 1800–10,600 ms, followed by the mask (presented for 484 ms), a prime (i.e., a neutral or threatening picture) that was presented for 16 ms, and again the mask (presented for 100 ms). Subsequently, an unfamiliar-looking, a familiar-looking or a scrambled face was presented for 2000 ms. Thus, there were six conditions: a familiar-looking face presented after a neutral prime (neutral-familiar), a familiar-looking face presented after a threatening prime (threat-familiar), an unfamiliar-looking face presented after a neutral prime (neutral-unfamiliar), an unfamiliar-looking face presented after a threatening prime (threat-unfamiliar), a scrambled face presented after a neutral prime (neutral-scrambled), and a scrambled face presented after a threatening prime (threat-scrambled). Stimulus sequences (mask-prime-mask-[scrambled] face) were presented in quasi-random order, with the restriction that the same prime could not be presented more than twice in a row, the same face could not be repeated more than four times in a row, and the same condition could not repeat more than twice. In all, 13 neutral and 13 threatening primes were each presented three times with each face, resulting in 39 (3 × 13) trials per condition. To ensure that participants remained alert during the task, they had to press a button in order to continue the task after every 11–13 trials. The average duration of the task was 23 min.

Primes

The stimuli used as primes were developed by Nummenmaa, Hirvonen, Parkkola, & Hetanen (2008). To enable comparability between neutral and threatening primes, these authors created pairs of photographs depicting a neutral and a threatening scene, respectively. Each pair was matched on luminosity, global energy, contrast density, and complexity, and showed the same persons in comparable proximity to each other. Each photograph portrayed two persons. On threatening photographs, interpersonal attack scenes (e.g., one person strangling the other) were shown, whereas non-emotional situations (e.g., two persons having a conversation) were depicted on neutral photographs.

We selected 13 pairs out of the 37 pairs of threatening and neutral pictures (Nummenmaa et al., 2008): an independent sample of 15 participants were presented with the pictures for 16 ms, with forward and backward masking as described above, and asked to press one button if they were sure a neutral picture had been presented, a second button if they were sure a threatening picture had been presented (they were instructed to press these buttons only if they had seen the picture and were sure of its contents), and a third button if they had not seen the picture or were unsure of its contents. This was done to test whether the neutral and threatening pictures were visible for the participants when these pictures were presented for 16 ms. Ideally, the participants should not be able to consciously perceive and identify the pictures, since our goal was to investigate subliminal processing of neutral and threatening stimuli. Therefore, only pictures that were not identified as neutral or threatening above chance levels (i.e., pictures for which significantly more than 50% of participants answered "unsure") were selected for use in the current study. Another independent sample of 28 participants was used to rate the 13 pairs of pictures for valence and arousal. Threatening photographs ($M = 8.40$, $SD = 0.22$) were rated as significantly more negative than neutral photographs ($M = 4.48$, $SD = 0.60$; $t_{(12)} = -23.90$, $p < 0.01$, $d = -8.67$), on a scale ranging from 1 ("positive") to 9 ("negative"). Moreover, on a scale ranging from 1 ("affected") to 9 ("calm"), threatening primes ($M = 3.43$, $SD = 0.41$) evoked significantly more arousal than neutral primes ($M = 7.31$, $SD = 0.33$; $t_{(12)} = 21.62$, $p < 0.01$, $d = 10.43$).

At the end of the second session, participants in the current study were asked whether they had seen any of the pictures presented in between the masks (i.e., the primes). Twenty-six participants (58%) indicated that they had noticed the pictures. Subsequently, these participants were asked to indicate which of several items (e.g., "truck", "adults") they had seen in the pictures. Some of these items had actually been present in the pictures, others had not. None of the participants performed above chance level, the participants selected seen and unseen items with equal probability.

Facial Stimuli

Pictures of unfamiliar- and familiar-looking children were created by morphing the photograph of a child's face (unfamiliar to the participant) with: (i) a photograph of an unknown female's face and (ii) a photograph of the participant's own face. Prior to the first session, participants were asked to provide a full-color digital photograph of themselves that met the following criteria: picture on a light and uniform background, showing their face (full frontal) and neck only, with a neutral facial expression, and no piercings, make-up or glasses. Full color, full frontal photographs of two female faces (both Caucasian and unfamiliar to the participant, aged 24 and 25 year, neutral facial expression, no jewelry or glasses) were used to create the unfamiliar-looking morphs. For half of the participants, female face 1 was used to create the unfamiliar-looking morph for session one and female face 2 was used to create the unfamiliar-looking morph for session two, and for the other half vice versa. Full color, full frontal photographs of six 9–11 year old children (three boys and three girls, all Caucasian [but slightly varying in skin color], all unfamiliar to the participants, with neutral facial expression, no jewelry or glasses) were available for morphing. For half the participants ($n = 21$ for the current sample) morphs were created with the picture of a female child and for the other half ($n = 24$ for the current sample) morphs were created with the picture of a male child. Within genders, the child that best matched the participant's skin color and face-shape was selected for ease of morphing. Both unfamiliar-looking and familiar-looking morphs were created with the photograph of the same child. One familiar-looking and two-unfamiliar-looking morphs were created for the two sessions. We did not use the same unfamiliar-morph for both sessions, since this would have led to increased familiarity with the unfamiliar-looking face in session two compared to session one.

Prior to morphing, all photographs were resized to 448 × 560 pixels and edited using Adobe Photoshop CS: External features (i.e., hair and ears) were removed and the pictures were framed on a black background. Morphing was then performed using Fantamorph 5 Deluxe, such that the picture of the familiar-looking child consisted for 50% of the participant's face and for 50% of an unknown child's face, and the picture of the unfamiliar-looking child consisted for 50% of the unknown female's face and for 50% of the child's face. The resulting pictures appear to present children slightly older than the 9–11 year olds used for morphing. An independent sample of 15 participants rated the age of the unfamiliar-looking morphs as 13.80 years (*SD* = 1.66) and the familiar-looking morphs as 14.40 years (*SD* = 1.60) on average ($p > 0.05$).[1] Finally, a scrambled face was created for each participant from the familiar-looking morph by randomly rearranging blocks of 9 × 9 pixels using Matlab R2012B.

At the end of the second session, participants in the current study evaluated how much the familiar-looking and unfamiliar-looking faces used during the priming task resembled themselves on a scale ranging from 0% resemblance to 100% resemblance. On average, the participants reported a similarity of 38.07% (*SD* = 13.38%) with the familiar and 6.40% (*SD* = 6.84%) with the unfamiliar morphs.[2] The difference in perceived similarity was significant with a large effect size ($t_{(44)} = 15.82$, $p < 0.01$, $d = 2.98$).

Image Acquisition

Images were acquired at the Leiden University Medical Center on a 3-T Philips Achieva MRI system (Philips Medical Systems, Best, Netherlands) with a 32-channel SENSE (Sensitivity Encoding) head coil. An event-related design with 680 T2*-weighted whole-brain echo planar images (EPI, repetition time (TR) = 2200 ms, echo time (TE) = 30 ms., flip angle = 80°, 38 transverse slices, descending acquisition order, voxelsize = 2.75 × 2.75 × 3.025 mm^3 with a 10% interslice gap, field of view (FOV) = 220 × 114.675 × 220 mm^3) was used for the functional scans. To avoid magnetic saturation effects, the first four functional scans were discarded. In addition, an anatomical 3D T1-weighted scan (TR = 9.825 ms, TE = 4.605 ms, flip angle = 8°, 140 transverse slices, voxelsize 0.875 × 0.875 × 1.2 mm^3, FOV = 224 × 168 × 177.333 mm^3) and a high-resolution T2*-weighted EPI-image (TR = 2200 ms, TE = 30 ms, flip angle = 80°, 84 transverse slices, voxel size = 1.964 × 1.964 × 2 mm^3, FOV = 220 × 168 × 220 mm^3) were obtained for coregistration purposes.

---

[1] One participant's scores were swapped. The correct ratings are: M= 13.87 years (SD = 1.73) for the unfamiliar-looking morphs, and M= 14.33 years (SD = 1.54) for the familiar-looking morphs (p > 0.05).

[2] In converting VAS scores to percentages resemblance we erroneously divided scores by 10 instead of 6. The correct estimates of the similarity ratings are M= 63.46% (SD= 22.30%) for the familiar and M=10.67% (SD= 11.40%) for the unfamiliar morphs.

fMRI Data Analysis

Data-analyses were performed using FSL (FMRIB's Software Library[3]) FEAT (FMRI Expert Analysis Tool) version 5.0.4, part of Jenkinson, Beckmann, Behrens, Woolrich, & Smith (2012) and Smith et al. (2004). The following pre-statistics processing steps were carried out: motion correction using MCFLIRT (Jenkinson et al., 2002), non-brain removal (BET; Smith, 2002), spatial smoothing using a Gaussian kernel with a full-width-at-half-maximum of 6 mm, and high-pass temporal filtering with a high-pass filter cutoff of 100 s.

Functional images were registered to the high-resolution EPI-image, which was then registered to the 3D T1-weighted scan, and then to the 2 mm isotropic MNI-152 standard space image (T1 standard brain averaged over 152 subjects; Montreal Neurological Institute, Montreal, QC, Canada; Jenkinson, Bannister, Brady, & Smith 2002). Functional activity in response to the stimuli was investigated using general linear model analysis in native space. Because primes and masks were presented for very short durations and time-locked to the presentation of the faces, hemodynamic responses to the individual stimuli within a mask-prime-mask-face sequence overlap extensively and sum to a total, summed hemodynamic response to the stimulus sequence. Assuming that responses to the masks in a given brain area do not vary systematically across the different conditions (as these are defined by different types of primes and the faces, but the masks are always the same), this summed response may vary depending on the response to the primes and faces. We thus treated the presentation of mask-prime-mask-face as a single stimulation period, and thus the different conditions (threat-familiar, threat-unfamiliar, threat-scrambled, neutral-familiar, neutral-unfamiliar, and neutral-scrambled) and the participants' responses were modeled as seven explanatory variables using the Custom (3 column format) wave function and convolved with a double gamma hemodynamic response function. The temporal derivatives of the explanatory variables were included in the model, yielding 14 regressors. Subsequently, individual lower-level contrast images (see below) were submitted to higher-level mixed effects (FLAME 1 + 2) group ROI and whole-brain analyses. Group means for ROIs and whole-brain analyses were tested using F-tests. All statistical images were thresholded using clusters determined by $Z > 2.3$

---

[3] www.fmrib.ox.ac.uk/fsl

(F-values are automatically converted to z-statistics) and a cluster-corrected significance threshold of $p < 0.05$ (Worsley, 2001)[4].

Before evaluating our main hypotheses, a preliminary analysis was conducted to check whether faces activated known face processing areas such as the fusiform gyrus more than scrambled stimuli. For this purpose, the contrast face (i.e., neutral-familiar, neutral-unfamiliar, threat-familiar, threat-unfamiliar) > scrambled face was tested. In the preliminary analysis, no confound regressors or continuous predictors were added to the model and only whole-brain analysis was conducted. Results of the preliminary analysis, showing that the facial stimuli reliably activated face processing areas as expected, can be found in the Supplementary material.

To evaluate our main hypotheses, separate whole-brain and ROI-analyses were performed to test for: (i) differences in brain activity in response to stimulus sequences in which faces were presented with a neutral prime and sequences in which faces were presented with a threatening prime; (ii) differences in brain activity in response to familiar and unfamiliar faces; and (iii) interactions between the type of face and the type of prime. For these analyses, five contrasts of interest were calculated: (1) familiar (threat-familiar and neutral-familiar) vs. unfamiliar (threat-unfamiliar and neutral-unfamiliar); (2) threatening (threat-familiar and threat-unfamiliar) vs. neutral (neutral-familiar and neutral-unfamiliar); (3) (threat-familiar vs. neutral-familiar) vs. (threat-unfamiliar vs. neutral-unfamiliar); (4) threat-familiar vs. neutral-familiar; and (5) threat-unfamiliar vs. neutral-unfamiliar. The first contrast

---

[4] Due to the large number of voxels analyzed, multiple testing is a well-known problem inherent to fMRI research. We chose to use cluster-extent based correction to correct for multiple testing. This correction procedure combines a threshold for results at individual voxels (i.e., for a cluster of voxels to be considered significantly activated all z-values at individual voxels within that cluster must exceed a certain value, in this case 2.3; t- and F-*values* are automatically converted to z-values) with requirements for the size of the cluster (i.e., a cluster is only considered significant if it consists of a sufficient number of voxels), in such a way that the probability of finding active clusters under the null hypothesis is smaller than 0.05 (i.e., α). To offer some indication of the robustness of statistical findings it is common to report the z- and p-values (α) employed in the cluster-based correction procedure, as well as the cluster size (in number of voxels) of significant clusters and the maximum z-value (i.e., Z-max) found among individual voxels in each significant cluster.

tested for differences in brain activity in response to viewing familiar-looking faces compared to unfamiliar-looking faces. Because the type of face presented may be expected to affect only the hemodynamic response to the face stimulus (as the prime is presented before it), the areas identified respond differently to familiar and unfamiliar faces. The second contrast tested for effects of the primes, i.e., differences in brain activity in response to presentation of sequences including neutral primes compared to sequences including threatening primes. Because the type of prime presented could theoretically affect the hemodynamic response to both the prime itself and the face stimulus, this contrast will identify both brain regions that respond differentially to the neutral and threatening primes (i.e., areas involved in processing the primes) and brain areas that respond differently to faces (regardless of whether this was a familiar or unfamiliar face) depending on the type of prime (i.e., a priming effect on face processing). In case of significant effects, comparisons to sequences including a scrambled stimulus instead of a face are used to distinguish between these two options. The third contrast tested for the interaction (i.e., variation in the effect of familiarity depending on the type of prime and/or variation in the effect of priming depending on face familiarity), and significant results for contrasts 4 and 5 were only interpreted in areas where contrast 3 was significant. F-tests were used to evaluate the hypotheses of the whole-brain and ROI-analyses. Scores on love withdrawal and empathic concern were included as continuous predictors and handedness was added to the model as a confound regressor.

The ROI analyses were performed on bilateral amygdala, bilateral inferior frontal gyrus (IFG) and bilateral STG to test our *a priori* hypotheses. Three higher-level analyses, restricted to bilateral amygdala, IFG and STG respectively, were conducted to investigate activity in these regions with maximized statistical power by limiting the number of statistical tests to the investigated ROI. The Harvard-Oxford Subcortical Structures Atlas was used to define the ROI for the amygdala and the Harvard-Oxford Cortical Structures Atlas (both implemented in FSL version 5.0.4) was used to define ROIs for the IFG and the STG. Three masks were created in 2 mm isotropic MNI-152 standard space (Jenkinson et al., 2002), consisting of voxels belonging to the left or right amygdala, IFG and STG respectively with a probability of at least 25%. Exploratory whole-brain analyses were performed to investigate brain activity in regions other than the *a priori* ROIs.

As use of caffeine may have an influence on brain activity as measured with fMRI (Liu et al., 2004; Liau, Perthen, & Liu, 2008; Perthen, Lansing, Liau, Liu, & Buxton 2008; Chen & Parrish, 2009), we reran the ROI and whole brain analyses testing effects of face and prime type excluding participants ($n = 7$) who did not comply with the request to abstain from caffeine during the last 12 h before the study. Using this sensitivity analysis, we evaluated whether effects in the total sample were replicated in the sample without caffeine-using respondents.

Results

All significant clusters were defined by $Z > 2.3$ and a cluster-corrected significance threshold of $p < 0.05$ (Worsley, 2001).

ROI Analyses

Significant effects of face familiarity were found in both the IFG and STG (see Figure 1): familiar-looking faces elicited greater brain activity than unfamiliar-looking faces in the right IFG (size = 220, *Z-max* = 4.54, MNI coordinates *x, y, z* (mm) = 46, 26, 22), whereas unfamiliar-looking faces elicited greater brain activity than familiar-looking faces in bilateral STG (cluster 1 [left]: size = 304, *Z-max* = 4.1, MNI coordinates *x, y, z* (mm) = −62, −32, 14, cluster 2 [right]: size = 182, *Z-max* = 3.53, MNI coordinates *x, y, z* (mm) = 64, −26, 10). No effects of familiarity were found in the amygdala, and we did not find significant activity differences between stimuli preceded by threatening and neutral primes or any familiarity*prime type interaction in any of the ROIs. Love withdrawal and empathic concern did not affect brain activity in any of the ROIs either.

Whole-Brain Analyses

To explore effects of threat priming and familiarity in regions outside our regions of interest, we performed whole-brain analyses. Results of these analyses revealed more widespread effects of face familiarity on brain activity (see Table 1). As illustrated in Figure 2, familiar-looking faces elicited greater brain activity in a cluster including not only the right IFG, but also parts of the right middle frontal gyrus (MFG), frontal pole, and insular cortex (cluster 2: size = 1559, peak *Z-max* = 4.54, MNI coordinates *x, y, z* (mm) = 46, 26, 22; see

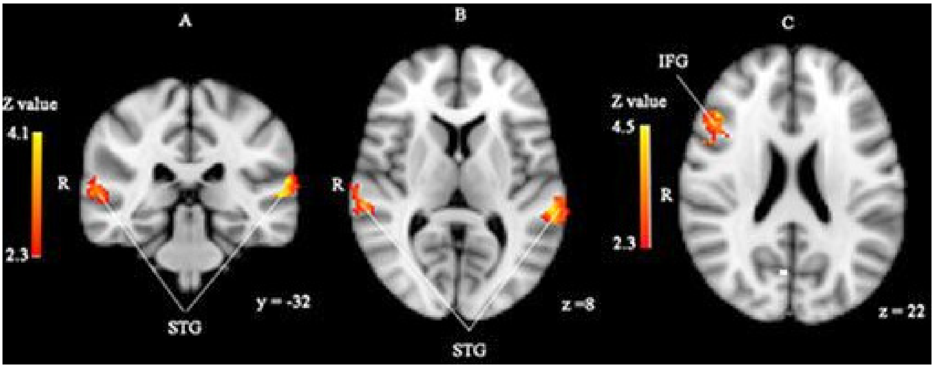


Figure 1. Significantly enhanced activity in bilateral superior temporal gyrus (STG) in response to unfamiliar compared to familiar faces in the sample with 45 participants (A, B). Significantly enhanced activity in right inferior frontal gyrus (IFG) in response to familiar compared to unfamiliar faces (C). ROI analyses, $p < 0.05$, corrected by cluster threshold ($Z > 2.3$).

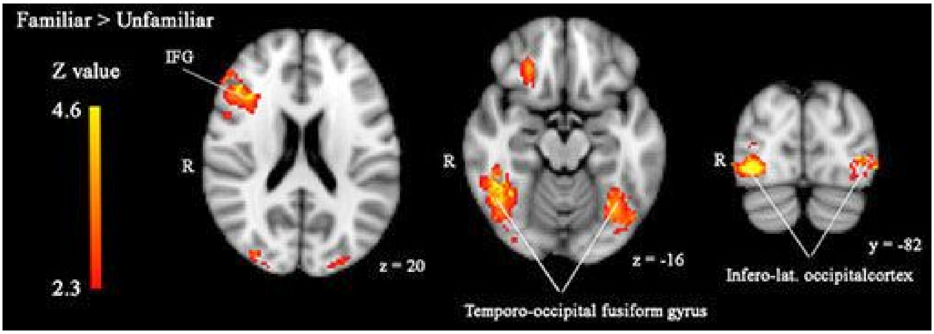


Figure 2. Significantly enhanced activity in right IFG, bilateral temporo-occipital fusiform gyrus, and infero-lateral occipital cortex in response to familiar compared to unfamiliar faces. Whole brain analyses, $p < 0.05$, corrected by threshold ($Z > 2.3$).

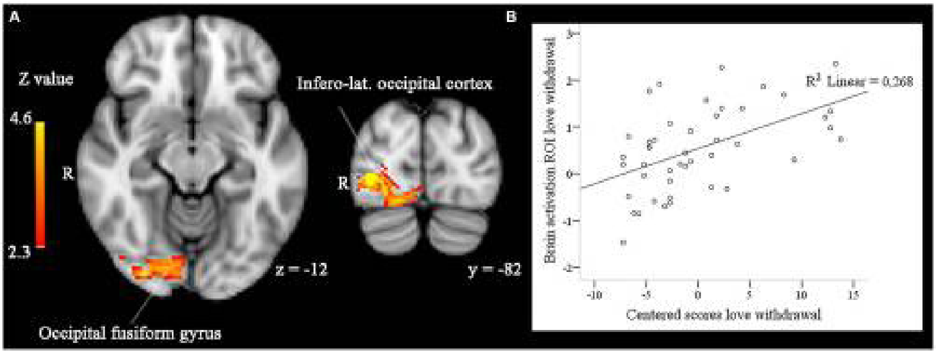


Figure 3.(A) In right infero-lateral occipital cortex and right occipital fusiform gyrus the effect of face familiarity (enhanced actvivity to familiar compared to unfamiliar faces) is significantly and positively related to participants' scores on love withdrawal in whole-brain analyses, $p < 0.05$, corrected by cluster threshold ($Z > 2.3$). (B) Scatterplot between the activity difference (familiar > unfamiliar) found in these areas and participants' centered scores on love withdrawal.

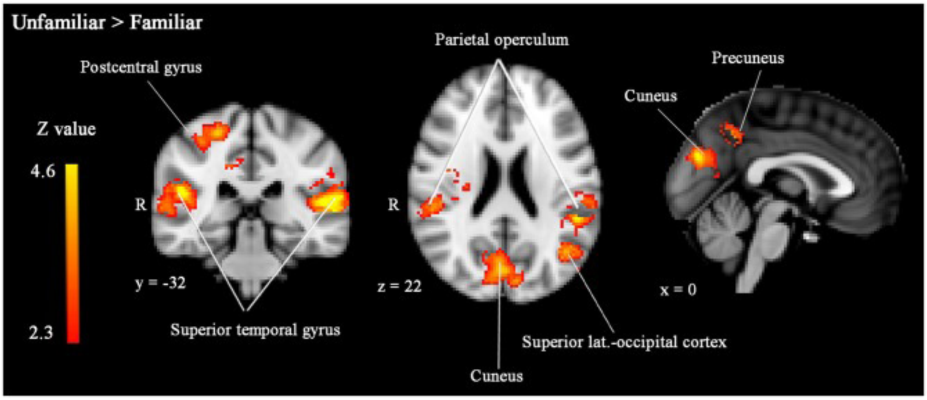


Figure 4. Significantly enhanced activity in bilateral STG, right postcentral gyrus, bilateral parietal operculum, left superior lateral occipital cortex, bilateral cuneus and left precuneus in reponse to unfamiliar compared to familiar faces. Whole-brain analyses, $p < 0.05$, corrected by cluster threshold ($Z > 2.3$).

Figure 2). We also found increased activity in response to familiar- compared to unfamiliar-looking faces in bilateral clusters including the occipital pole, infero-lateral occipital cortex, and the temporo-occipital fusiform gyrus (cluster 3 [right]: size = 1627, peak *Z-max* = 4.63, MNI coordinates *x, y, z* (mm) = 40, −72, −10, cluster 1 [left]: size = 1182, peak *Z-max* = 4.29, MNI coordinates *x, y, z* (mm) = −30, −90, 6).

In addition, love withdrawal interacted with the effect of familiarity in a partially overlapping cluster including the right infero-lateral occipital cortex, occipital fusiform gyrus, and occipital pole (size = 1008, peak *Z-max* = 4.22, MNI coordinates *x, y, z* (mm) = 34, −80, 0). As illustrated in Figure 3, the effect of familiarity was larger (i.e., a larger difference in brain activity in response to familiar-looking compared to unfamiliar-looking faces) for participants reporting more love withdrawal.

Table 1. MNI Coordinates and *Z*-max values for regions with significant main effects for face familiarity and for regions in which love withdrawal and empathic concern interact with effects of face familiarity.

Experimental effect	Clusters	Size	Region	*Z-max*	MNI coordinates for Z-max		
					x	y	z
Familiar > unfamiliar	3	1627	Right infero-lateral occipital cortex	4.63	40	-72	-10
	2	1559	Right MFG	4.54	46	26	22
	1	1182	Left infero-lateral occipital cortex	4.29	-30	-90	6
Unfamiliar > familiar	4	2815	Left planum temporale	4.57	-56	-30	10
	3	1794	Right postcentral gyrus	3.86	24	-34	60
	2	1504	Right planum temporale	4.09	48	-30	16
	1	765	Left cuneus	3.85	0	-78	26
Familiar > unfamiliar $^{LW+}$	1	1008	Right infero-lateral occipital cortex	4.22	34	-80	0

Unfamiliar-looking faces compared to familiar-looking faces evoked increased activity bilaterally in clusters including not only the STG, but also the posterior division of the supramarginal gyrus, and the parietal operculum, and extending anteriorly into the planum temporale (cluster 4 [left]: size = 2815, peak *Z-max* = 4.57, MNI coordinates *x, y, z* (mm) = −56, −30, 10; cluster 2 [right]: size = 1504, peak *Z-max* = 4.09, MNI coordinates *x, y, z* (mm) = 48, −30, 16). In addition, unfamiliar-looking faces compared to familiar-looking faces elicited heightened activity in a cluster including the right postcentral gyrus, right superior parietal lobe, and bilateral precuneus (cluster 3: size = 1794, peak *Z-max* = 3.86, MNI coordinates *x, y, z* (mm) = 24, −34, 60) and in bilateral cuneus (cluster 1: size = 765, peak *Z-max* = 3.85, MNI coordinates *x, y, z* (mm) = 0, −78, 26). These clusters are presented in Figure 4.

Empathic concern interacted with face familiarity in a frontal area (frontal pole: size = 593, peak *Z-max* = 3.68, MNI coordinates *x, y, z* (mm) = −22, 56, 36) far at the outside of the brain and not overlapping with any of the significant clusters described above, suggesting artifactual activity.

Similar to the ROI analyses, the whole-brain analyses did not reveal any significant activity differences between stimuli preceded by threatening and neutral primes or any familiarity*prime type interaction.

Analyses with 38 Participants

The clusters that we found in the ROI and whole-brain analyses in the total sample of 45 participants were largely replicated with the 38 participants who abstained from caffeine use. In the smaller sample, however, familiar-looking compared to unfamiliar-looking faces evoked enhanced activity only in right occipital pole, infero-lateral occipital cortex, and temporo-occipital fusiform gyrus and not bilaterally as in the sample with 45 participants. This small difference may be due to lower statistical power in the smaller sample.

Discussion

In this study, we investigated the processing of subliminally presented threatening primes and their effects on neural responses to pictures of a familiar (and potentially "own") and unfamiliar child in a homogenous student sample. In addition, we studied moderating effects of empathy and experiences of love-withdrawal. Since we were particularly interested in parental protective reactions in the presence a potential threat, we combined pictures of faces of familiar-looking children with primes that depicted threatening scenes. Contrary to our expectations, we did not find evidence of a priming effect, nor of any interaction between empathy or experienced love withdrawal and priming. The primes used in our study depicted fairly complex neutral and threatening scenes, showing multiple people and objects. It is possible that these images were too complex for the brain to extract the threatening or neutral content fast and efficiently, and that preconscious processing of the threat thus did not occur. Earlier studies on subliminal priming usually used less complex stimuli (e.g., fearful or angry facial expressions; e.g., Morris et al., 1998; Whalen et al., 1998; Almeida et al., 2013). However, if the brain possesses a specialized threat-detection system to enable fast and automatic responses to environmental threats, we would expect such a system to be able to process threatening stimuli with various contents and complexities.

An absence of priming effects on face processing may not only be due to the complexity of the scenes used as primes. It is also possible that the scenes did not induce protective reactions. It seems natural, however, that parents or adults in general look after children when they witness or expect threatening events and their own survival is not immediately at stake. Although caregiving responses may be weaker when the child is not in distress (i.e., crying) or when the threat is aimed at the adult, the hypothesis that neural differences can be observed if the subliminal threat is really processed may still be warranted. Interestingly, results of some recent studies actually do not provide much support for the existence of a human brain system capable of preconscious processing of threatening information (see Pessoa & Adolphs, 2010; Hoffmann et al., 2012). It should be noted that in earlier studies of "subliminal" processing of affective information primes were often presented for 30 ms or longer (Morris et al., 1998; Whalen et al., 1998; Dimberg, Thunberg, & Elmehed, 2000; Li, Zinbarg, Boehm, & Paller, 2008). Participants differ in their sensitivity to threatening stimuli, but reliable detection of fearful faces has been observed with presentation durations of only 17 ms (Pessoa, Japee, & Ungerleider 2005). This suggests that priming may not have been completely subliminal in the previous experiments with prime presentations of approximately 30 ms. Obviously, what is needed is replication of our study with subliminal stimuli of varying duration and involving a more direct threat to the child or, alternatively, with supraliminal threat stimuli.

We also investigated the effects of face familiarity on neural activity. As hypothesized, we found enhanced activity in response to familiar-looking faces in the IFG, extending into the MFG and insular cortex. Enhanced activity in IFG and MFG is frequently seen in familiar face processing (Gobbini & Haxby, 2006; Platek & Kemp, 2009; Taylor et al., 2009). In addition, enhanced activity of these brain areas is frequently found in response to pictures or videos of an own vs. other/unfamiliar child (Bartels & Zeki, 2004; Noriuchi, Kikuchi, & Senoo, 2008; Kuo, Carp, Light, & Grewen, 2012; Wittfoth-Schardt et al., 2012). These effects may be associated with the role of these brain areas in self-referential processing: when confronted with (the face of) someone who physically resembles the self (whether due to kinship or otherwise) concepts relating to the self are automatically activated. This human tendency to extrapolate from physical, "outer", resemblance to psychological, "inner", resemblance plays an important role in the understanding (including empathic understanding) of others (see for a review Devue & Brédart, 2011) Importantly, these as well as other processes in which the IFG and insula have an important role (e.g., emotion-regulation) are very important for parental behavior. In fact, Swain et al. (2014) have given these areas an important role in their model of the "parental brain". We also found enhanced activity in occipital and temporal (i.e., occipital pole, infero-lateral occipital cortex, and fusiform gyrus) areas involved in visual, and, more specifically, face processing (Haxby, Hoffman, & Gobbini, 2000; Natu & O'Toole, 2011). Our findings suggest preferential processing of the familiar-looking faces. Both of these effects fit well with known processing advantages of stimuli associated with own compared to unfamiliar children in parents (Leibenluft et al., 2004).

Love withdrawal moderated the effect of familiarity in right hemisphere face processing areas (infero-lateral occipital cortex and occipital fusiform gyrus). Participants reporting more love withdrawal showed larger differences in brain activity in response to familiar-looking vs. unfamiliar-looking faces. Interestingly, changes in the neural processing of facial stimuli in young adults reporting high maternal love withdrawal have been observed before (e.g., Huffmeijer et al., 2011). Experiences of love withdrawal create a mental link between behavior and relational consequences, and they compromise the security of the parent-child attachment relationship, which becomes conditional on the child's behavior (Goldstein & Heaven, 2000; Assor et al., 2004; Elliot & Thrash, 2004; Renk et al., 2006). We suggest that the increased salience and relevance of the parent-child relationship may generalize to relationships more generally and increase the processing of information relevant to those relationships, in particular relationships with other family members, including own children. The enhanced brain activity seen in participants with high scores on love withdrawal to familiar-looking faces, designed to appeal to a kinship bond, may be a neural signature of this processing enhancement.

In contrast to familiar-looking faces, unfamiliar-looking faces enhanced activity in bilateral STG, and in whole brain analyses this activity extended anteriorly from the planum temporale into the parietal operculum and the posterior part of the supramarginal gyrus. In addition, unfamiliar-looking faces elicited enhanced activity in the right postcentral gyrus, right superior parietal lobe, and bilateral cuneus and precuneus. These regions are part of the brain's socio-emotional networks and they are, in particular the superior temporal sulcus (STS), involved in ToM processes. ToM refers to the cognitive capacity to attribute mental states (e.g., desires, intentions) to others and to predict others' behaviors from these mental states (Frith & Frith, 1999; Schurz, Radau, Aichhorn, Richlan, & Perner, 2014). Although contrasting results exist in the literature (Leibenluft et al., 2004), several previous studies have observed decreased activity in the STS in response to familiar faces (Ramon et al., 2015) and to pictures of mothers' own children (Bartels & Zeki, 2004) compared to unfamiliar faces. Decreased activity in brain areas supporting ToM in response to familiar compared to unfamiliar faces may be explained by reduced effort, i.e., due to for example self-referential processing (see above) it is easier to estimate the mental state of someone familiar or similar to the self and by a lower need to investigate the social validity, i.e., a reduced need to thoroughly assess/estimate the mental state or intentions of familiar persons, as suggested by Bartels & Zeki (2004).

Future research should also take some limitations of the current study into account. The most important limitation is of course the use of morphed faces instead of faces of own offspring. The difference between the "own" and unfamiliar children's faces was physical resemblance (looking familiar). Although the participants reported afterwards that the familiar faces were much more similar to their own faces than the unfamiliar faces and physical resemblance is a kinship cue, replication with faces of real offspring is needed to disentangle effects of biological relatedness and familiarity or physical resemblance in the absence of a kinship bond. Second, the current design did not allow for separate modeling of hemodynamic responses to primes and faces. Because primes and masks were presented for very short durations and time-locked to the presentation of the faces, hemodynamic responses to the individual stimuli within a mask-prime-mask-face sequence overlapped extensively, requiring the modeling of a single, summed hemodynamic response to each stimulus sequence. Although all relevant processes (processing of the primes, processing of the faces, and effects of priming on face processing) could be separated by comparing responses to the different stimulus sequence conditions, it is worthwhile to consider the inclusion of conditions in which either the prime or the face is omitted from the stimulus sequence. Although this would lengthen the paradigm, such a design would allow for separate modeling of responses to primes and faces which may lead to a less complicated analysis approach. Third, we used self-report questionnaires to investigate parental love withdrawal and empathy. There are obvious limitations to the accuracy and reliability of participants' self-reports. Furthermore, we chose to focus our analyses on the affective component of empathy captured by the Empathic Concern scale of the Interpersonal Reactivity Index (IRI). Future studies may also focus on other empathy dimensions, such as the cognitive component or the tendency to experience personal distress. In addition, the participants in our sample reported relatively low levels of experiences of love withdrawal and were generally psychologically healthy. It may be interesting to replicate the study within clinical samples, e.g., in parents reporting experiences of (emotional) abuse or those with post-traumatic stress symptoms. These experiences and symptoms have been related to hyper-vigilance and arousal (van Harmelen et al., 2013; Stark et al., 2015), and individuals with post-traumatic symptoms in particular seem to have enhanced amygdala responses to threat (Stark et al., 2015). Thus, they may be more sensitive to (supraliminally or subliminally presented) threat primes, which may lead to altered priming effects. Third, as this was a pilot study focusing on women without children, for which pictures of an "own" child were artificially created, future studies should certainly focus on parents' neural responses to their own and unfamiliar children in the presence and absence of threat. As we found no evidence of subliminal priming effects, replication studies might want to modify the priming design. Less complex stimuli (e.g., angry or fearful vs. neutral faces) could be used, as the neutral and threatening primes used in our study were perhaps too complex. Primes could also be presented supraliminally. In fact, ideally the primes should be both sub- and supraliminally, to directly compare brain activity seen in both conditions and to shed light on the possibility of subliminal threat processing. Finally, a behavioral measure (e.g., reaction time) of the priming effect could be included to directly compare changes in brain activity to changes in behavior.

So far, our results question the effectiveness of subliminal threat-priming. As others have questioned the existence of a fast and automatic threat processing system, too (see Pessoa & Adolphs, 2010; Hoffmann et al., 2012), we feel that this is an important issue that deserves attention in future research. In addition, our results again illustrate the profound impact that experienced parenting strategies such as love withdrawal may have, even at the level of neural information processing. Although changes in neural activity and the preferential processing of familiar vs. unfamiliar faces are not inherently adaptive and desirable or maladaptive and undesirable (as this ultimately depends on the characteristics and demands of the situation or context in which they occur), parental use of love withdrawal may generally be considered undesirable because of its behavioral consequences (see e.g., Assor et al., 2004; Renk et al., 2006). It certainly has to be considered insensitive parental behavior that elevates the chance for an insecure attachment relationship to develop (Bowlby, 1973/1985). The fact that even relatively "mild" negative parenting experiences, such as the levels of love withdrawal reported by our participants, are associated with changes in very basic neural processes only adds to the importance of the early parenting environment and attachment relationship for individual development.

References

Almeida, J., Pajtas, P. E., Mahon, B. Z., Nakayama, K., and Caramazza, A. (2013). Affect of the unconscious: Visually suppressed angry faces modulate our decisions. *Cogn. Affect. Behav. Neurosci.* 13, 94–101. doi: 10.3758/s13415-012-0133-7

Assor, A., Roth, G., and Deci, E.L. (2004). The emotional costs of parents' conditional regard: a self-determination theory analysis. *J. Pers.* 72, 47–88. doi: 10.1111/j.0022-3506.2004.00256.x

Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., Riem, M. M. E., Tops, M., and Alink, L. R. A. (2012). Oxytocin decreases handgrip force in reaction to infant crying in females without harsh parenting experiences. *Soc. Cogn. Affect. Neurosci.* 7, 951–957. doi: 10.1093/scan/nsr067

Bakermans-Kranenburg, M. J. & Van IJzendoorn, M. H. (2017). Oxytocin and Human Sensitive and Protective Parenting. *Current topics in behavioral neurosciences.*

Bartels, A., & Zeki, M. (2004). The neural correlates of maternal and romantic love. *Neuroimage* 21, 1155–1166. doi: 10.1016/j.neuroimage.2003.11.003

Beyers, W., and Goossens, L. (2003). Psychological separation and adjustment to university: moderating effects of gender, age and perceived parenting style. *J. Adolesc. Res.* 18, 363–382. doi: 10.1177/0743558403018004003

Bowlby, J. (1973/1985). "Attachment and loss", in *Separation: Anxiety and Anger*, Vol. 2. Harmondsworth: Penguin Books.

Bowlby, J. (1988). "Attachment, communication and the therapeutic process", in *A Secure Base: Parent-Child Attachment and Healthy Human Development*, (London: Routledge), 137–157.

Bressan, P., and Grassi, M. (2004). Parental resemblance in 1-years-olds and the Gaussian curve. *Evol. Hum. Behav.* 25, 133–141. doi: 10.1016/j.evolhumbehav.2004.03.001

Chen, Y., and Parrish, T.B. (2009). Caffeine dose effect on activation-induced BOLD and CBF responses. *Neuroimage* 46, 577–583. doi: 10.1016/j.neuroimage.2009.03.012

Davis, M. H. (1983). Measuring individual differences in empathy: evidence for a multidimensional approach. *J. Pers. Soc. Psychol.* 44, 113–126. doi: 10.1037/0022-3514.44.1.113

DeBruine, L.M. (2004). Resemblance to self increases the appeal of child faces to both men and women. *Evol. Hum. Behav.* 25, 142–154. doi: 10.1016/j.evolhumbehav.2004.03.003

De Corte, K., Buysse, A., Verhofstadt, L. L., Roeyers, H., Ponnet, K., and Davis, M. H. (2007). Measuring empathic tendencies: reliability and validity of the Dutch version of the interpersonal reactivity index. *Psychol. Belg.* 47, 235–260.

Devue, C., and Brédart, S. (2011). The neural correlates of visual self-recognition. *Conscious. Cogn.* 20, 40–51. doi: 10.1016/j.concog.2010.09.007

Dimberg, U., Thunberg, M., and Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychol. Sci.* 11, 86–89. doi: 10.1111/1467-9280.00221

Elliot, A. J., and Thrash, T. M. (2004). The intergenerational transmission of fear of failure. *Pers. Soc. Psychol. Bull.* 30, 957–971. doi: 10.1037/e519342007-001

Eres, R., Decety, J., Louis, W. R., and Molenberghs, P. (2015). Individual differences in local gray matter density are associated with differences in affective and cognitive empathy. *Neuroimage* 117, 305–310. doi: 10.1016/j.neuroimage.2015.05.038

Euser, E. M., van IJzendoorn, M. H., and Prinzie, P., and Bakermans-Kranenburg, M. J. (2010). Prevalence of child maltreatment in the Netherlands. *Child Maltreat.* 15, 5–17. doi: 10.1177/1077559509345904

Frith, C. D., and Frith, U. (1999). Interacting minds: a biological basis. *Science* 286, 1692–1695. doi: 10.1126/science.286.5445.1692

George, C., and Solomon, J. (2008). "The caregiving system: a behavioral systems approach to parenting", in *Handbook of Attachment: Theory Research and Clinical Applications*, 2nd Edn, eds J. Cassidy and P. R. Shaver (New York, NY: Guilford Press), 833–856.

Globisch, J., Hamm, A. O., Esteves, F., and Ohman, A. (1999). Fear appears fast: temporal course of startle reflex potentiation in animal fearful subjects. *Psychophysiology* 36, 66–75. doi: 10.1017/s0048577299970634

Gobbini, M. I., and Haxby, J. V. (2006). Neural response to the visual familiarity of faces. *Brain Res. Bull.* 71, 76–82. doi: 10.1016/j.brainresbull.2006.08.003

Goldstein, M., and Heaven, P. L. (2000). Perceptions of the family, delinquency and emotional adjustment among youth. *Pers. Individ. Dif.* 29, 1169–1178. doi: 10.1016/s0191-8869(99)00264-0

Haxby, J. V., Hoffman, E. A., and Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends Cogn Sci.* 4, 223–233. doi: 10.1016/s1364-6613(00)01482-0

Hoffmann, M., Lipka, J., Mothes-Lasch, M., Miltner, W. R., and Straube, T. (2012). Awareness modulates response of the amygdala and the visual cortex to highly arousing visual threat. *Neuroimage* 62, 1439–1444. doi: 10.1016/j.neuroimage.2012.05.064

Hoffman, M. L. and Saltzstein, H. D. (1967). Parent discipline and the child's moral development. *J. Pers. Soc. Psychol.* 5, 45–57.

Huffmeijer, R., Alink, L. R. A., Tops, M., Grewen, K. M., Light, K. C., Bakermans-Kranenburg, M. J., et al. (2013). The impact of oxytocin administration and maternal love withdrawal on event-related potentials (ERP) responses to emotional faces with performance feedback. *Horm. Behav.*, 63, 399–410. doi: 10.1016/j.yhbeh.2012.11.008

Huffmeijer, R., Tops, M., Alink, L. R. A., Bakermans-Kranenburg, M. J., and Van IJzendoorn, M. H. (2011). Love withdrawal is related to heightened processing of faces with emotional expressions and incongruent emotional feedback: evidence from ERPs. *Biol. Psychol.* 86, 307–313. doi: 10.1016/j.biopsycho.2011.01.003

Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17, 825–841. doi: 10.1006/nimg.2002.1132

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., and Smith, S. M. (2012). FSL. *Neuroimage* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09.015

Kilner, J. M., Neal, A., Weiskopf, N., Friston, K. J., and Frith, C. D. (2009). Evidence of mirror neurons in human inferior frontal gyrus. *J. Neurosci.*, 29, 10153–10159. doi: 10.1523/jneurosci.2668-09.2009

Klimecki, O., and Singer, T. (2013). "Empathy from the perspective of social neuroscience", in *The Crambridge Handbook of Human Affective Neuroscience*, eds J. Armony and P. Vuillemier (New York, NY: Cambridge University Press), 533–549.

Koch, C., and Tsuchiya, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends Cogn. Sci.* 11, 16–22. doi: 10.1016/j.tics.2006.10.012

Kuo, P. X., Carp, J., Light, K. C., and Grewen, K. M. (2012). Neural responses to infants linked with behavioral interactions and testosterone in fathers. *Biol. Psychol.* 91, 302–306. doi: 10.1016/j.biopsycho.2012.08.002

LeDoux, J. (1998). Fear and the brain: where have we been, and where are we going? *Biol. Psychiatry* 44, 1229–1238. doi: 10.1016/s0006-3223(98)00282-0

Leibenluft, E., Gobbini, M. I., Harrison, T., and Haxby, J. V. (2004). Mothers' neural activation in response to pictures of their children and other children. *Biol. Psychiatry* 56, 225–232. doi: 10.1016/j.biopsych.2004.05.017

Li, W., Zinbarg, R. E., Boehm, S. G., and Paller, K. A. (2008). Neural and behavioral evidence for affective priming from unconsciously perceived emotional facial expressions and the influence of trait anxiety. *J. Cogn. Neurosci.* 20, 95–107. doi: 10.1162/jocn.2008.20006

Liau, J., Perthen, J. E., and Liu, T. T. (2008). Caffeine reduces the activation extent and contrast-to-noise ratio of the functional cerebral blood flow response but not the BOLD response. *Neuroimage* 42, 296–305. doi: 10.1016/j.neuroimage.2008.04.177

Lieberman, M. D., and Eisenberger, N. I. (2015). The dorsal anterior cingulate cortex is selective for pain: Results from large-scale reverse inference. *Proc. Natl. Acad. Sci. USA* 112, 15250–15255. doi: 10.1073/pnas.1515083112

Liu, T. T., Behzadi, Y., Restom, K., Uludag, K., Lu, K., Buracas, G. T., et al. (2004). Caffeine alters the temporal dynamics of the visual BOLD response. *Neuroimage* 23, 1402–1413. doi: 10.1016/j.neuroimage.2004.07.061

Locke, L. M., and Prinz, R. J. (2002). Measurement of parental discipline and nurturance. *Clin. Psychol. Rev.* 22, 895–929. doi: 10.1016/s0272-7358(02)00133-2

Maloney, L. T., and Dal Martello, M. F. (2006). Kin recognition and the perceived facial similarity of children. *J. Vis.*, 6, 1047–1056. doi: 10.1016/j.visres.2008.09.025

Masten, C. L., Morelli, S. A., and Eisenberger, N. I. (2011). An fMRI investigation of empathy for 'social pain' and subsequent prosocial behavior. *Neuroimage* 55, 381–388. doi: 10.1016/j.neuroimage.2010.11.060

Mikulincer, M., Shaver, P. R., Gillath, O., and Nitzberg, R. A. (2005). Attachment, caregiving and altruism: boosting attachment security increases compassion and helping. *J. Pers. Soc. Psychol.* 89, 817–839. doi: 10.1037/0022-3514.89.5.817

Morris, J. S., Öhman, A., and Dolan, R. J. (1998). Conscious and unconscious emotional learning in the human amygdala. *Nature* 393, 467–470. doi: 10.1038/30976

Natu, V., and O'Toole, A. J. (2011). The neural processing of familiar and unfamiliar faces: A review and synopsis. *Br. J. Psychol.* 102, 726–747. doi: 10.1111/j.2044-8295.2011.02053.x

Noriuchi, M., Kikuchi, Y., and Senoo, A. (2008). The functional neuroanatomy of maternal love: Mother's response to infant's attachment behaviors. *Biol. Psychiatry* 63, 415–423. doi: 10.1016/j.biopsych.2007.05.018

Nummenmaa, L., Hirvonen, J., Parkkola, R., and Hietanen, J. K. (2008). Is emotional contagion special? An fMRI study on neural systems for affective and cognitive empathy. *Neuroimage* 43, 571–580. doi: 10.1016/j.neuroimage.2008.08.014

Patrick, R. B., and Gibbs, J. C. (2007). Parental expression of disappointment: should it be a factor in Hoffman's model of parental discipline? *J. Genet. Psychol.* 168, 131–145. doi: 10.3200/gntp.168.2.131-146

Perthen, J. E., Lansing, A. E., Liau, J., Liu, T. T., and Buxton, R. B. (2008). Caffeine induced uncoupling of cerebral blood flow and oxygen metabolism: A calibrated-BOLD fMRI study. *Neuroimage* 40, 237–247. doi: 10.1016/j.neuroimage.2007.10.049

Pessoa, L., and Adolphs, R. (2010). Emotion processing and the amygdala: from a 'low road to 'many roads' of evaluation biological significance. *Nat. Rev. Neurosci.* 11, 773–783. doi: 10.1038/nrn2920

Pessoa, L., Japee, S., Sturman, D., and Ungerleider, L. G. (2006). Target visibility and visual awareness modulate amygdala responses to fearful faces. *Cereb. Cortex* 16, 366–375. doi: 10.1093/cercor/bhi115

Pessoa, L., Japee, S., and Ungerleider, L. G. (2005). Visual awareness and the detection of fearful faces. *Emotion* 5, 243–247. doi: 10.1037/1528-3542.5.2.243

Platek, S. M., and Kemp, S. M. (2009). Is family special to the brain? An event-related fMRI study of familiar, familial, and self-face recognition. *Neuropsychologia* 47, 849–858. doi: 10.1016/j.neuropsychologia.2008.12.027

Platek, S. M., Raines, D. M., Gallup, G. J., Mohamed, F. B., Thomson, J. W., and Myers, T. E., et al. (2004). Reactions to children's faces: males are more affected by resemblance than females are and so are their brains. *Evol. Hum. Behav.* 25, 394–405. doi: 10.1016/j.evolhumbehav.2004.08.007

Platek, S. M., Wathne, K., Tierny, N. G., and Thomson, J. W. (2008). Neural correlates of self-face recognition: an effect-location meta-analysis. *Brain Res.* 1232, 173–184. doi: 10.1016/j.brainres.2008.07.010

Psychology Software Tools, Inc. [E-Prime 2.0]. (2012). Available online at: http://www.pstnet.com

Ramon, M., Vizioli, L., Liu-Shang, J., and Rossien, B. (2015). Neural microgenesis of personally familiar face recognition. *Proc. Natl. Acad. Sci. USA* 112, E4835–E4844. doi: 10.1073/pnas.1414929112

Renk, K., McKinney, C., Klein, J., and Oliveros, A. (2006). Childhood discipline, perceptions of parents and current functioning in female college students. *J. Adolesc.* 29, 73–88. doi: 10.1016/j.adolescence.2005.01.006

Schludermann, E. H., and Schludermann, S. M. (1988). *Children's Report of Parent Behavior (CRPBI-108, CRPBI-30) for Older Children and Adolescents.* (Technical Report). MB, Canada: University of Manitoba.

Schludermann, S., and Schludermann, E. (1983). Sociocultural change and adolescents perceptions of parent behavior. *Dev. Psychol.* 19, 674–685. doi: 10.1037/0012-1649.19.5.674

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., and Perner, J. (2014). Fractional theory of mind: a meta-analysis of functional brain imaging studies. *Neurosci. Biobehav. Rev.* 42, 429–434. doi: 10.1016/j.neubiorev.2014.01.009

Shamay-Tsoory, S. G. (2011). The neural base for empathy. *Neuroscientist* 17, 18–24, doi: 10.1177/1073858410379268

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., and Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science* 303, 1157–1162. doi: 10.1126/science.1093535

Smith, S. M. (2002). Fast robust automated brain extraction. *Hum. Brain Mapp.* 17 143–155. doi: 10.1002/hbm.10062

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 23, S208–219. doi: 10.1016/j.neuroimage.2004.07.051

Stark, E. A., Parsons, C. E., Van Hartevelt, T. J., Charquero-Ballester, M., McManners, H., Ehlers, A., et al. (2015). Post-traumatic stress influences the brain even in the absence of symptoms: a systematic, quantitative meta-analysis of neuroimaging studies. *Neurosci. Behav. Rev.* 56, 207–221. doi: 10.1016/j.neubiorev.2015.07.007

Strathearn, L., Fonagy, P., Amico, J., and Montague, P. R. (2009). Adult attachment predicts maternal brain and oxytocin response to infant cues. *Neuropsychopharmacology* 34, 2655–2666. doi: 10.1038/npp.2009.103

Swain, J. E., Kim, P., Spicer, J., Ho, S. S., Dayton, C. J., Elmadih, A., et al. (2014). Approaching the biology of human parental attachment: Brain imaging, oxytocin and coordinated assessments of mothers and fathers. *Brain Res.* 1580, 78–101. doi: 10.1016/j.brainres.2014.03.007

Tabachnick, B. G., and Fidell, L. S. (2001). *Using Multivariate Statistics,* 4th Edn. Boston: Allyn and Bacon.

Taylor, M. J., Arsalidou, M., Bayless, S. J., Morris, D., Evans, J. W. and Barbeau, E. J. (2009). Neural correlates of personally familiar faces: parents, partner and own faces. *Hum. Brain Mapp.* 30, 2008–2020. doi: 10.1002/hbm.20646 van Harmelen, A., van Tol, M., Demenescu, L. R., van der Wee, N. A., Veltman, D. J. and Aleman, A., et al. (2013). Enhanced amygdala reactivity to emotional faces in adults reporting childhood emotional maltreatment. *Soc. Cogn. Affec. Neurosci.* 8, 362–369. doi: 10.1093/scan/nss007

Van IJzendoorn, M. H., Huffmeijer, R., Alink, L. R. A., Bakermans-Kranenburg, M. J. and Tops, M. (2011). The impact of oxytocin administration on charitable donating is moderated by experiences of parental love-withdrawal. *Front. Psychol.* 2:258. doi: 10.3389/fpsyg.2011.00258

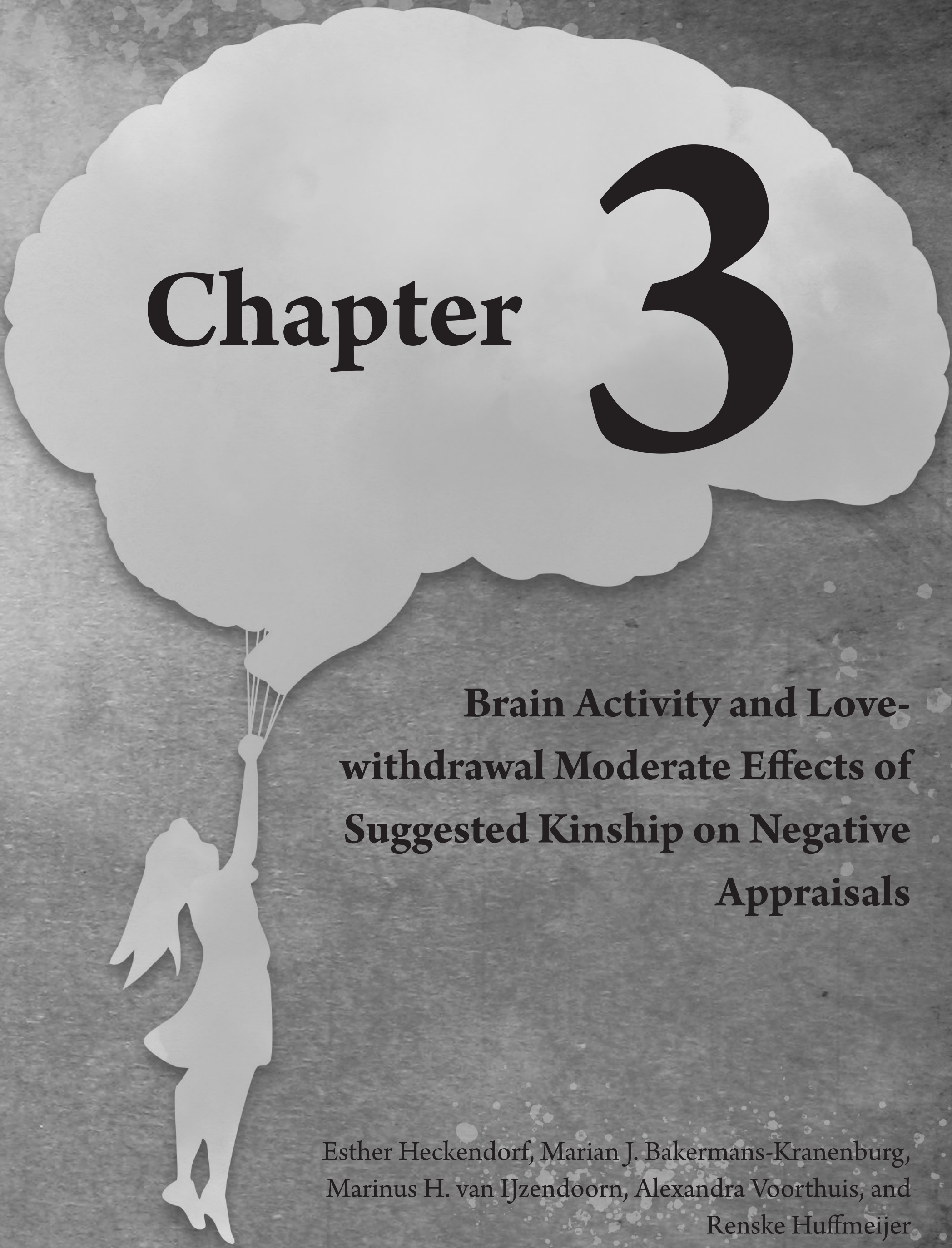
van Strien, J. W. (1992). Classificatie van links- en rechtshandige proefpersonen. *Nederlands Tijdschrift voor de Psychologie en Haar Grensgebieden* 47, 88–92.

van Strien, J. W. (2003). *Vragenlijst voor Handvoorkeur.* Available online at http://hdl.handle.net/1765/955

Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., and Jenike, M. A. (1998). Masked presentation of emotional facial expressions modulate amygdala activity without explicit knowledge. *J. Neurosci.* 18, 411–418.

Wittfoth-Schardt, D., Gründing, J., Wittfoth, M., Lanfermann, H., Heinrichs, M., Domes, G., et al. (2012). Oxytocin modulates neural reactivity to children's faces as a function of social salience. *Neuropsychopharmacology* 37, 1799–1807. doi: 10.1038/npp.2012.47

Worsley, K. J. (2001). "Statistical analysis of activity images", in *Functional MRI: An Introduction to Methods,* eds P. Jezzard, P. M. Matthews, and S. M. Smith (Oxford: Oxford University Press), 251–270.



Chapter 3

Brain Activity and Love-withdrawal Moderate Effects of Suggested Kinship on Negative Appraisals

Esther Heckendorf, Marian J. Bakermans-Kranenburg,
Marinus H. van IJzendoorn, Alexandra Voorthuis, and
Renske Huffmeijer

Evolutionary Behavioral Sciences (in press)

Abstract

Facial resemblance serves as an important kinship cue in humans, and, as such, facilitates kin recognition. Mechanisms to facilitate kin recognition exist in many different species and have probably evolved to promote nepotism and avoid inbreeding. Responses to facial resemblance may however be affected by a person's own (childhood) experiences with close relatives. In the present study, we investigated whether the degree of resemblance of children's faces with the participant's face was related to participants' positive and negative appraisals of the children's faces. We morphed pictures of an unfamiliar child's face with the participant's face and with the face of an unfamiliar adult, to create facial stimuli that differed in their degree of facial resemblance with the participant. We examined the effects of childhood experiences with parental love-withdrawal and participants' neural processing of facial identity (fusiform face area activity, FFA). As hypothesized, negative appraisal of the faces decreased linearly with increasing facial resemblance. In addition, love-withdrawal and FFA activity moderated the relation between facial resemblance and negative appraisal. Participants who both reported high love-withdrawal and showed greater FFA activity showed the largest decrease in negative appraisals with increasing resemblance. Positive appraisal of the faces was not associated with resemblance of the child face with the participant's face. Future research should address the effects of phenotypic kinship cues on actual parental behavior.

Introduction

How do we recognize our relatives? Various cues, like co-residence during childhood, observing your mother caring for a newborn sibling, and phenotypic cues of genetic relatedness influence our estimates of kinship (Lieberman, Tooby, & Cosmides, 2007; Porter & Moore, 1981; DeBruine, 2002; DeBruine, 2004a). An individual may use such cues to distinguish kin from non-kin (i.e. estimate whether he or she is related to another individual), and to determine the degree of genetic relatedness (i.e. determine how closely he or she is related to another individual). Many species are known to engage in kin recognition, and kin recognition mechanisms have probably evolved to promote nepotism (i.e. favor relatives) and avoid inbreeding (Hepper, 2011).

According to the inclusive fitness theory, nepotistic behavior pays off because it promotes survival and reproductive success of close relatives who share a proportion of one's genes (Hamilton, 1964), and thus enhances the chance that an individual's genes are transmitted. Inbreeding, on the other hand, is related to negative outcomes, including increased childhood mortality rates and risk of genetic abnormalities in offspring (Al-Gazali, Hamamy, & Al-Arrayad, 2006; Saha, Hamad & Mohamed, 1990). Therefore, avoiding inbreeding and promoting nepotism are highly desirable in evolutionary terms, and mechanisms to facilitate kin recognition are adaptive because they enable individuals to adjust their behavior according to their genetic relatedness with others (Lehmann & Perrin, 2002). Phenotype matching, i.e. the process of determining the degree of relatedness to another individual by comparing that individual's phenotypic cues (e.g. odor, auditory signals) to a 'kin protoype', has been identified as a crucial kinship recognition mechanism in many animals (Hauber & Sherman, 2001).

In humans, facial resemblance may play an important role in phenotype matching, as it has been shown that individuals are able to make fairly accurate judgments of others' actual relatedness based on facial similarity (Bressan & Grassi, 2004; Maloney & Dal Martello, 2006; Kaminski, Dridi, Graff, & Gentaz 2009; Alvergne, Faurie & Raymond, 2010). In addition, participants rated adults' faces that resembled themselves higher on prosocial characteristics such as trustworthiness (DeBruine, 2002, 2005), and participants were willing to show more prosocial behavior (e.g. cooperation) to adults that facially resembled themselves (Krupp, DeBruine & Barclay, 2008). Likewise, adults were more willing to invest in children whose faces resembled their own (Bressan, Bertamini, Nalli, & Zanutto, 2009; DeBruine, 2004b; Platek, Burch, Panyavin, Wasserman, & Gallup 2002). Lastly, facial resemblance enhances the perceived attractiveness of children's faces (DeBruine, 2004b), particularly for same-sex faces (DeBruine, 2004a). Interestingly, the effects of facial resemblance of other-sex faces was found to be sensitive to the relational context: In one study, facial resemblance of other-sex faces was not related to perceived attractiveness in the context of a long-term relationship (in which both prosocial regard and sexual attraction are expected to play a role). However, when presented in the context of a short-term relationship (in which primarily sexual attraction is expected to play a major role), facial resemblance of other-sex faces was related to decreased attractiveness (DeBruine, 2005). Taken together, findings from previous studies thus fit well with the theory that kin recognition serves the dual goal of both promoting nepotism and avoiding inbreeding.

However, individual differences may exist in the way people respond to individuals that resemble themselves. It can be expected that effects of facial resemblance (and kinship cues in general) depend on an individual's own (childhood) experiences with close relatives. We recently found that young adults' experiences with parental love-withdrawal during childhood were related to face processing. Parental love-withdrawal was related to increased brain activity in face processing areas (i.e. infero-lateral occipital cortex and fusiform areas) as well as increased activity in the inferior frontal gyrus in response to children's faces that resembled their own compared to children's faces that did not resemble their own face (Heckendorf, Huffmeijer, Bakermans-Kranenburg, & Van IJzendoorn, 2016). Thus, experiences of love-withdrawal were related to enhanced neural differentiation between children's faces that resembled the participants' face and children's faces that did not. Love-withdrawal is a parental disciplinary strategy in which the parent withholds love and affection when the child misbehaves or disobeys, and is associated with long-term negative consequences, including low emotional well-being, low self-esteem, fear of failure, insecure attachment, and feelings of resentment toward the parents (Bowlby, 1973/1985, p. 243; Assor, Roth, & Deci, 2004; Goldstein & Heaven, 2000; Elliot & Thrash, 2004; Renk, McKinney, Klein, & Oliveros 2006). Previous research has also associated experiences of love-withdrawal with enhanced (neural) processing of and responding to interpersonal cues, including emotional faces, and with differential effects of experimental manipulations such as oxytocin administration on these processes (Bakermans-Kranenburg, Van IJzendoorn, Riem, Tops, & Alink, 2012; Huffmeijer, Tops, Alink, Bakermans-Kranenburg, & Van IJzendoorn, 2011; Huffmeijer et al., 2013; Riem et al., 2013; Van IJzendoorn, Huffmeijer, Alink, Bakermans-Kranenburg, & Tops, 2011). Thus, the regular use of love-withdrawal as a parenting technique seems to have a profound impact on an individual's long-term psychological well-being and interpersonal relationships (including kinship and romantic bonds), as well as on the neural processing of relevant stimuli. Such effects may well be reflected in a different evaluation of self-resembling faces.

Individual variation in the evaluation of faces may result from individual differences in the neural processing of faces. A large body of research has robustly and consistently related face processing to activity in the fusiform face area (FFA), a brain region located on the lateral side of the middle fusiform gyrus (see Kanwisher & Yovel, 2006 for a review). Important with respect to facial resemblance as a kinship cue is that FFA activity has specifically been related to analyzing facial identity: the particular features and configuration that define an individual face (Andrews & Ewbank, 2004; Andrews & Schluppeck, 2004; Grill-Spector, Knouf, & Kanwisher, 2004; Yovel & Kanwisher, 2005). Greater FFA activity may thus be considered to reflect more extensive processing of the observed face, in particular the identifying features and configuration. This may in turn impact on the perception and differentiation of facial resemblance and thus affect subsequent evaluations.

In the present study, we examined whether the degree of self-resemblance of child faces affects participants' evaluations of the child faces. More specifically, we examined whether increasing resemblance of a child's face with the participant's face is related to more positive appraisals and less negative appraisals. In addition, we examined whether effects of facial resemblance depend on participants' neural processing of facial identity (FFA activity) and whether childhood experiences of parental love-withdrawal moderate participants' reactions to facial resemblance. As analysis of facial identity is necessary for differential responding to different individuals, we expected stronger increases in positive appraisals and more exaggerated decreases in negative appraisals with increasing facial resemblance in individuals who process faces more extensively (i.e., show greater FFA activity in response to faces in general). We also expected that individuals with more experiences of parental love-withdrawal would show larger increases for positive appraisals and stronger decreases for negative appraisals with increasing facial resemblance. Most pronounced effects were thus expected in individuals who both process faces relatively extensively and report higher levels of love-withdrawal experiences.

Materials and methods

Participants

Forty-six female undergraduate and graduate students aged 21.81 years on average (*SD* = 2.60; range 18-28 years) completed two experimental sessions, 4-12 weeks apart (*M* = 4.61, *SD* = 1.68 weeks). Exclusion criteria were MRI contraindications, pregnancy, current psychiatric and neurological disorders, history of severe head injury, current alcohol or drug abuse, and chronic use of medication (except contraceptives). One participant provided no usable fMRI data due to excessive head movement, and three participants fell asleep during fMRI data acquisition. Thus, our final sample consisted of 42 participants aged 18 to 28 years (*M* =21.83 years, SD = 2.64). The Ethics Committee of the Leiden University Medical Center approved the experiment and informed consent was obtained from all participants at the beginning of the first session. All participants were rewarded with 40€ for participation. The structural MRI scans of the final sample did not show any anomalies.

Procedure

Participants were instructed to abstain from alcohol and excessive physical activity during the last 24 hours and from caffeine during the last 12 hours before the start of each session. At the start of session 1, participants filled out questionnaires, including a measure of parental love-withdrawal. The researchers explained the MRI procedure to the participants at the beginning of each session. Inside the scanner, foam inserts were placed between the head coil and the participant's head to minimize head movements. The participants watched two types of children's faces (faces showing no resemblance and faces showing 50% resemblance with the participant), as well as scrambled faces (as a visual control) preceded by subliminal primes (for detailed information see Heckendorf et al., 2016) within the scanner. Stimuli were projected onto a screen placed outside the opening of the scanner bore. Participants viewed the screen through a mirror fixed to the head coil. At the end of the second session, participants completed a rating task in which they evaluated both the children's faces shown in the fMRI experiment and a child's face resembling the participant for 75% on several positive and negative characteristics. Subsequently, participants were debriefed about the nature of the priming task presented in the MRI scanner.

Facial Stimuli

We created pictures of children showing no facial resemblance, 50% facial resemblance, and 75% facial resemblance with the participant by morphing the photograph of a child's face (unfamiliar to the participant) with a photograph of: (i) an unknown female face and (ii) a photograph of the participant's own face. Prior to the first session, participants provided a full-color digital photograph of themselves that met the following criteria: a light and uniform background, the picture showing their face (full frontal) and neck only, with a neutral facial expression, without piercings, make-up, or glasses. To create morphs not resembling the participant, we used full color, full frontal photographs of two female faces (both Caucasian and unfamiliar to the participant, aged 24 and 25 years, neutral facial expression, no jewelry or glasses). For half of the participants, the non-resembling morph for session 1 was created using female face 1 and the non-resembling morph for session 2 was created using female face 2, and for the other half vice versa. Full color, full frontal photographs of six 9–11 year old children (three boys and three girls, all Caucasian but slightly varying in skin color, all unfamiliar to the participants, with neutral facial expression, no jewelry or glasses) were available for morphing. For about half of the participants (n = 21), the morphs were created with the picture of a female child, for the other half (n = 24) with the picture of a male child. Within genders, we selected the child that best matched the participant's skin color and face-shape for ease of morphing. The same child was used for both the non-resembling and the 50% and 75% self-resembling morphs. We used both 50% and 75% self-resembling morphs in the rating task to investigate whether participants' ratings of the morphs would be linearly affected by self-resemblance.

Prior to morphing, all photographs were resized to 448 × 560 pixels and edited using Adobe Photoshop CS: External features (i.e., hair and ears) were removed and the pictures were framed on a black background. Subsequently, we performed morphing using Fantamorph 5 Deluxe. The 75%-resembling morph consisted for 75% of the participant's face and for 25% of an unknown child face, the 50%-resembling morphs consisted for 50% of the participant's face and for 50% of the same child face, and the non-resembling morphs consisted for 50% of the unknown female face and for 50% of the same child face. The 75%- and 50%-resembling morphs thus differed only in the relative degree to which they consisted of the participant's face (50% vs. 75%). Finally, we generated a scrambled face for each participant from the 50%-resembling morph by randomly rearranging blocks of 9 × 9 pixels using Matlab R2012B. These scrambled stimuli were used as a visual control in fMRI recording. An example of the morphing results is shown in Figure 1.

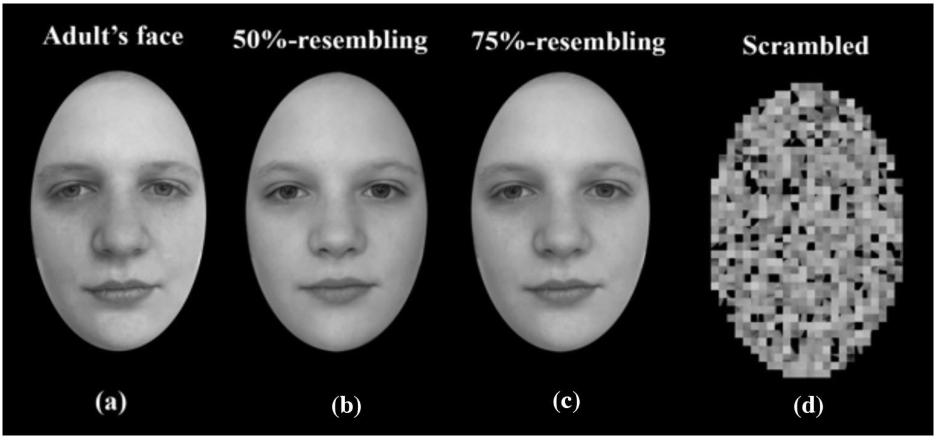


Figure 1. The 50%-resembling morph (b) consisted for 50% of the child's face and for 50% of the participant's face (a). The 75%-resembling morphs (c) consisted for 25% of the child's face and for 75% of the participant's face (a). The scrambled face (d) was generated by randomly rearranging pixels from the 50%-resembling morph (b).

An independent sample (N=15) rated the age of all morphed pictures used in our experiment. The estimated age of the three types of morphs (non-resembling, 50%-resembling and 75%-resembling) differed significantly ($F_{(2, 28)} = 15.72$, $p<.001$), as tested with a repeated measures ANOVA. Post-hoc pairwise comparisons with Bonferroni correction for multiple comparisons revealed that the 75%-resembling morph was perceived as older than both the non-resembling morph ($t_{(14)} = 2.13$, $p = .001$) and the 50%-resembling morph ($t_{(14)} = -1.67$, $p< .003$). Differences in age estimates between the 50% -resembling and the non-resembling morph ($p> .05$) were not significant.

Face Rating Task

At the end of the second session, participants completed a computerized task to evaluate the two non-resembling morphs, the 50%-resembling morph, and the 75%-resembling morph. E-Prime Software (Psychology Software Tools, 2012) was used to implement the task and to collect the participants' responses. The task consisted of three blocks during which the participants were asked to rate the morphs on several positive and negative characteristics. In the first block, participants were asked to indicate how attractive, intelligent, unreliable and worthwhile (Dutch: 'aantrekkelijk', 'slim', 'onbetrouwbaar' and 'de moeite waard') the different morphs were on a 600-point visual analog scale (VAS) ranging from 'not at all…' to 'very…' (e.g. ranging from 'not at all attractive' to 'very attractive').In the second block, participants rated how much commitment, connectedness, distance and aversion (Dutch: 'verbondenheid', 'betrokkenheid', 'afstandelijkheid' and 'afkeer') they felt toward the morphs, again on a 600-point VAS, ranging from 'no …. at all' to 'a lot of …' (e.g. ranging from 'no commitment at all' to 'a lot of commitment'). In the third block, participants evaluated how much the different morphs resembled themselves, on a 600-point VAS ranging from 0% to 100%. Each combination of morph and question was presented once. Thus, the first and the second block each consisted of 16 trials, and the third block consisted of 4 trials.

The blocks were presented in fixed order. Within a block, combinations of faces and questions were presented in quasi-random order, with the restrictions that the same face could not be presented more than two times in a row and that the 75%-resembling morph could not be presented on the very first trial of the task. On each trial, a morph (sized 6 x 9 cm) was presented in the center of the screen on a black background. A VAS was positioned below the morph. Participants used the mouse to rate the morphs by moving a slider to the position on the scale corresponding to their judgment. Participants could alter their choice until satisfied with their rating, and then moved to the next trial by clicking a button labeled 'next' that was presented in the bottom center of the screen. Only the participant's final choice on each trial was stored. Scores for the two non-resembling morphs were averaged.

Perceived self-resemblance of the morphs increased significantly with the degree of facial resemblance, with a large effect size ($F_{(2, 86)} = 357.69$, $p < .001$, $\eta_p^2 = .89$). Post-hoc comparisons with Bonferroni correction for multiple comparisons revealed that all differenceswere significant ($p < .001$). Thus, perceived self-resemblance of the 75%-resembling morph was significantly higher than perceived self-resemblance of the 50%-resembling morph, which in turn was significantly higher than perceived self-resemblance of the non-resembling morphs.

Face Processing: Fusiform Face Area Activity

Details concerning MRI image acquisition and fMRI processing and analysis, along with results (observed activation as well as reliability of activity) have been reported elsewhere, (Heckendorf et al., 2016). We analyzed the fMRI data withFSL (FMRIB's Software Library) FEAT (FMRI Expert Analysis Tool) version 5.0.4, part of Jenkinson, Beckmann, Behrens, Woolrich, and Smith (2012), and Smith et al. (2004). For the current study, we included FFA activity in the contrast faces (i.e., non-resembling, 50%-resembling) > scrambled stimuli during the second session as a measure of face processing. We selected this contrast to examine whether and how an individual's degree of face processing impacts on how she subsequently evaluates the morphs. We obtained significant activity for this contrast within the fusiform gyrus as well as other brain areas in the first session. Preliminary whole-brain analysis (identical to those conducted on data from the first session; see Heckendorf et al., 2016) also revealed significant activity in the fusiform gyrus for the contrast faces vs. scrambled in the second session. We chose to focus on fMRI data obtained during the second session because participants completed the Face Rating Task at the end of the second session.

We obtained each participant's maximum activity value within the FFA as our measure of face processing: As the FFA is an area within the fusiform gyrus defined by its preferential responding to faces, we first defined an anatomical mask of the fusiform gyrus using the Harvard-Oxford Cortical Structures Atlas. A binarized, anatomical mask including only voxels belonging to the right or left fusiform gyrus with a probability of at least 25% was created in 2 mm isotropic MNI-152 standard space (Jenkinson, Bannister, Brady & Smith, 2002). Next, we created a functional mask of face processing areas using the probability map obtained for a localizer task in an earlier study (N=124) for the contrast faces vs. scenes (Engell & McCarthy, 2013). We binarized and thresholded this contrast image (only voxels with at least a 25% probability to be significantly activated in the faces vs. scenes contrast included) in 2 mm isotropic MNI-152 standard space (Jenkinson et al., 2002). Finally, we multiplied the functional and anatomical masks to obtain a mask including only voxels that both respond specifically to faces and are part of the fusiform gyrus: thus a mask of the FFA. Featquery was used to export each participant's maximum activity value (contrast parameter) within the FFA mask to IBM SPSS Statistics 23 for further analysis. Our decision to include maximum activity values was based on the reliability of FFA activity obtained for the priming task, which was higher for maximum activity values than for mean and median activity values (Heckendorf, Bakermans-Kranenburg, Van IJzendoorn, & Huffmeijer, manuscript submitted for publication).

To ensure that excluding the three participants who fell asleep during the fMRI data acquisition did not significantly influence the results, we reran the analyses with these three participants included. For two of the participants that fell asleep during fMRI data acquisition, fMRI data of session 1 was available, and we included this participants´ maximum activity value (contrast parameter) within the FFA mask during session 1 in the analyses. For the third participant, fMRI data of the first 156 trials (67% of the task) of session 2 was available. As we obtained similar reliability of FFA activity in the contrast faces vs. scrambled stimuli in previous analyses (Heckendorf et al., manuscript submitted for publication), we included this participant's FFA activity during the available trials in the analyses.

Love-withdrawal Experiences

At the beginning of the first session, participants completed a modified version of the 30-item Children's Report of Parental Behavior Inventory (CRPBI-30, Schludermann & Schludermann, 1983; Beyers & Goossens, 2003), containing the items of the Acceptance and Psychological Control scales from the original questionnaire and several extra items to measure love-withdrawal. The 11-item Love-withdrawal scale consists of all five items that constitute the Withdrawal of Relations subscale of the 108-item CRPBI (3 of which are also included in the Psychological Control scale of the CRPBI-30; Schludermann & Schludermann, 1983), two items that were adapted from this same questionnaire, and four items that were adapted from the Parental Discipline Questionnaire (Patrick & Gibbs, 2007). Participants rated how well each item described their mother and father separately (e.g., "My mother was a person who if I'd hurt her feelings, stopped talking to me until I please her again") on a 5-point Likert scale, ranging from ("not at all") to ("very well"). We averaged scores for maternal and paternal love-withdrawal. We winsorized the score of one outlier ($z = 3.61$; the new score was identical to the second-highest score of the sample plus the difference between the second-highest and next-highest score, see Tabachnick & Fidell, 2001). After winsorizing, the scores were normally distributed with an average score of 18.55 (*SD* = 6.01, *Mdn* = 16.5, range: 11.5-32.5). Earlier studies obtained adequate validity and reliability of the CRPBI and its subscales (Schludermann & Schludermann, 1983, 1988; Locke & Prinz, 2002). Internal consistency of the scale in the current sample was high (Cronbach's alpha = .91). The Love-withdrawal subscale as used in this study was implemented in earlier research on the consequences of maternal love-withdrawal in young adults (e.g., Huffmeijer et al., 2011).

Statistical Analyses

Statistical analyses were conducted using SPSS software, version 23. We first conducted three principal component analyses (PCAs) with Varimax rotation, for the non-resembling, 50%-resembling, and 75%-resembling morphs, to determine which of the positive and negative ratings clustered. The PCAs yielded the same two factors for all three morphs, one containing the positive items (i.e. attractive, intelligent, worthwhile, committed and connected), and the other containing the negative items (i.e. unreliable, distant, repulsive). This two-factor solution explained 59.83% (50%-resembling morph), 61.34% (75%-resembling morph), and 61.61% (0%-resembling morph) of variance respectively. Eigenvalues were 3.36 (50%-resembling morph), 3.38 (non-resembling morph) and 3.57 (75%-resembling morph) for the first factor and 1.34 (75%-resembling morph), 1.43 (50%-resembling morph) and 1.55 (non-resembling morph) for the second factor. Loadings of the ratings on the factors were between 0.60 and 0.83. We therefore averaged across participants' ratings of attractiveness, intelligence, worthwhileness, commitment and connectedness to obtain a score for positive appraisal and across ratings of unreliability, distance, and repulsion to obtain a score for negative appraisal of each type of morph.

To examine effects of facial resemblance, experiences of love-withdrawal, and neural face processing (FFA activity) on participants' ratings, we conducted two repeated measures ANCOVAs with negative appraisals and positive appraisals respectively as dependent variables, facial resemblance (no resemblance, 50% resemblance, 75% resemblance) as within-subjects factor and experiences of love-withdrawal and FFA activity as continuous predictors. In case of sphericity violations, we applied the Greenhouse Geisser correction. In case of a significant main effect of facial resemblance, pairwise comparisons (Bonferroni-corrected for multiple comparisons) were computed.

Results

Means and standard deviations of participants' negative and positive appraisals of unfamiliar faces and faces resembling the participant for 50% and 75% are presented in Table 1.

Table 1. Average ratings and standard deviations of the negative and the positive scale for all face stimuli: Participants rated the faces that resembled themselves more positive and less negative.

Child faces	Negative		Positive	
	M	*SD*	*M*	*SD*
Unfamiliar	203.33	92.07	287.41	87.45
50% familiar	178.02	94.63	333.29	94.10
75% familiar	145.47	106.65	365.34	97.71

Negative appraisals

The repeated measures ANCOVA revealed a significant main effect of facial resemblance ($F_{(2, 76)} = 3.73$, $p = .03$, $\eta_p^2 = .09$), reflecting a linear effect of facial resemblance ($F_{(1, 38)} = 4.32$, $p = .045$, $\eta_p^2 = .10$): Faces that resembled the participants to a greater extent were evaluated less negatively. Although a clear linear decrease in negative appraisals was observed (see Table 1), post hoc comparisons revealed that the participants rated the 75%-resembling morph significantly less negatively than the non-resembling morph ($p < .01$), but participants' ratings of the 50% self-resembling morph did not differ significantly from ratings of the other two morphs (*ps*> .05).

The significant main effect of facial resemblance was qualified by a significant two-way interaction between facial resemblance and love-withdrawal ($F_{(2,76)} = 3.92$, $p = .02$, $\eta_p^2 = .09$) as well as a significant three-way interaction between facial resemblance, love-withdrawal and FFA activity ($F_{(2,76)} = 4.44$, $p = .02$, η_p^2

= .11). The interaction between facial resemblance and FFA activity was not significant ($F_{(2,76)} = 2.74$, $p = .07$).

To visualize the interaction effects we divided the participants into groups reporting fewer and more experiences of love-withdrawal using a median split (*Mdn* = 16.5), and we conducted separate repeated measures ANCOVAs with facial resemblance and FFA activity as independent variables for each group. The interaction between facial resemblance and FFA activity was significant. For participants who had experienced high levels of love-withdrawal ($F_{(2,38)} = 4.32$, $p = .02$, $\eta_p^2 = .19$), but not for participants who had experienced low levels of love-withdrawal ($p > .05$). To further explore the interaction, we therefore divided the participants that had experienced high levels of love-withdrawal into two groups with high (N=11) or low (N=10) FFA activity using a median split, and conducted repeated measures ANOVAs with facial resemblance as within-subjects factor for each group separately. A significant effect of

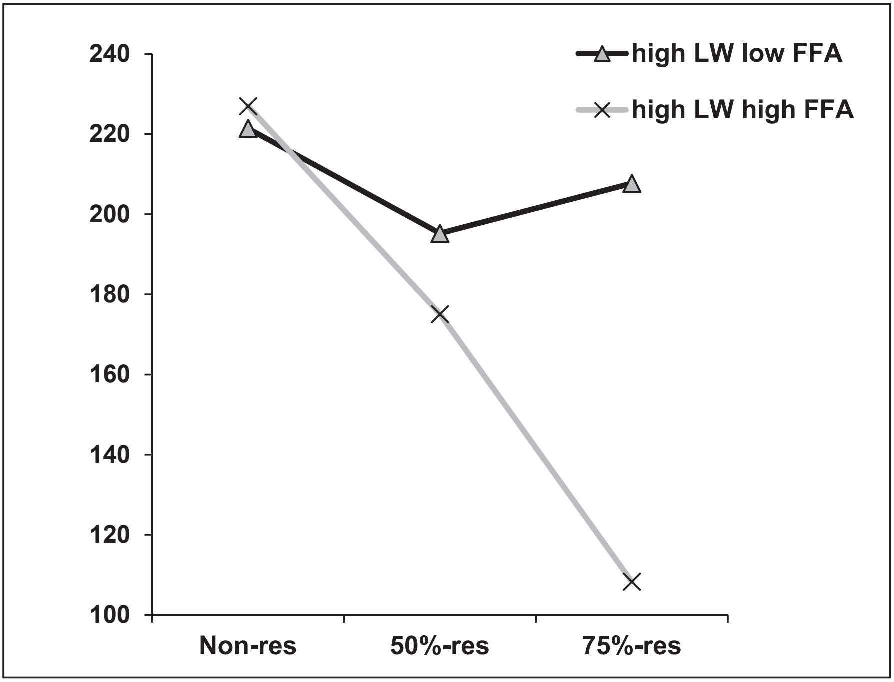


Figure 2. Interaction between facial resemblance and FFA activity in participants with high levels of reported love-withdrawal. Individuals with high levels of both FFA activity and experienced love-withdrawal showed the most pronounced decrease in negative appraisals.

facial resemblance was found only in the group showing high FFA activity ($F_{(2,20)}$ = 12.56, p = .00, η_p^2= .56; low FFA activity: $F_{(2,18)}$ = 0.27, p = .77, η_p^2= .03), reflecting a linear decrease in negative appraisals with increasing facial resemblance ($F_{(1,10)}$ = 19.19, p = .00, η_p^2= .66; see Figure 2). Thus, the effect of facial resemblance on negative appraisals of the morphs was most pronounced in individuals who had both experienced relatively much love-withdrawal and processed the faces to a large extent.

We reran the analyses with the bigger sample of 45 participants with incomplete fMRI data from session 2. These analyses revealed the same results as in the sample of 42 participants.

Positive Appraisals

Although, as shown in Figure 2, positive appraisals seemed to increase linearly with increasing facial resemblance, the repeated measures ANCOVA for positive appraisals revealed no significant main or interaction effects (all *F*s ≤ 1.65, *p*s> .20 [42 participants]; all *F*s ≤ 2.26, *p*s> .11 [45 participants]).

Discussion

We investigated young adults' appraisal of children's faces as a function of the degree of resemblance to their own face. As expected, increased self-resemblance was related to less negative appraisals of the children's faces. Furthermore, love-withdrawal and FFA activity moderated the relation between facial resemblance and negative appraisals, with more pronounced effects of facial resemblance in participants reporting high love-withdrawal and showing high FFA activity. Thus, particularly in participants who reported high levels of parental love-withdrawal, extensive face processing, as indicated by greater FFA activity, was associated with larger decreases in negative evaluations of children that resembled the participant.

The finding that participants showed less negative appraisals of child faces that resembled themselves more, implicating close kinship, makes sense from an evolutionary perspective: it may promote caregiving behaviors directed at closely related children, which in turn enhance the child's likelihood of survival and future reproductive success, thus increasing the chance that an individual's genes are transmitted to the following generation (Hamilton, 1963, 1964). The finding that children's faces that resemble the participant less are rated more negatively is also in line with the results of studies suggesting that both stepmothers and stepfathers invest less in their stepchildren than in their biological children (Anderson, Kaplan, & Lancaster, 1999; Anderson, Kaplan, Lam, & Lancaster, 1999; Evenhouse & Reilly, 2000; Hetherington & Jodl, 1994; Marlowe, 1999; Schmeeckle, 2007; Tifferet, Jorev, & Nasanovitz; 2010; Zvoch, 1999), and that stepchildren are more likely to be abused by their stepparents (Creighton & Noyes, 1989; Daly & Wilson, 1996; Van IJzendoorn, Euser, Prinzie, Juffer, & Bakermans-Kranenburg, 2009). In a Canadian study, the risk of 0-2 year-old children to be killed by a stepparent was estimated as even 70 times higher than the risk to be killed by a genetic parent (Daly & Wilson, 1988b). Similar results were obtained in a U.S. study, in which the estimated risk of fatal abuse by a stepparent in children under 3 years was about 100 times higher than the risk of fatal abuse by genetic parents (Daly & Wilson, 1988a). Stronger negative feelings towards biologically unrelated stepchildren may play a role in these increased risk rates among stepchildren, although results have not been unequivocal (Gelles & Harrop, 1991; Malkin & Lamb, 1994; Temrin, Buchmayer, & Enquist, 2000; but see Daly & Wilson, 2008).

The finding that a more pronounced decrease in negative appraisal with increasing facial resemblance was observed in participants reporting more experiences of love-withdrawal and showing greater FFA activity, indicate, first, that the extent to which individuals process faces varying in resemblance to the participant indeed affects their appraisal of these faces. The FFA analyzes face configuration in order to differentiate between individual faces (see for a review Kanwisher & Yovel, 2006), and greater FFA activity may reflect an enhanced ability to discriminate between identities and heightened sensitivity to identity cues, including facial resemblance. Individuals that process faces more extensively may therefore detect facial resemblance better, which may explain why participants showing higher FFA activity (in combination with experiences of love-withdrawal) show larger decreases of negative appraisals with increasing self-resemblance of children's faces.

The results also suggest that childhood experiences of parental love-withdrawal as a disciplinary strategy are related to changes in the neural processing of and reactions to facial resemblance in children's faces. The results are in line with previous findings relating experiences of love-withdrawal to enhanced effects of emotional expression on face processing (Huffmeijer et al.,

2011) and enhanced neural differentiation between self-resembling and non-self-resembling children (Heckendorf et al., 2016). Thus, regular experiences of parental love-withdrawal as a disciplinary strategy may not only be associated with long-term negative outcomes such as low self-esteem and reduced emotional well-being (Assor et al., 2004; Goldstein & Heaven, 2000; Renk et al., 2006), but also with changes in the neural processing of and reactions to interpersonal stimuli, including facial stimuli communicating emotions or kinship cues, that are still evident in young adulthood.

Contrary to our expectations and previous findings (DeBruine, 2002, 2003, 2004b, 2005; Krupp et al., 2008; Platek et al., 2002), effects were restricted to negative appraisals and did not extend to positive appraisals. Several factors may have contributed to such discrepancies. First, we studied female participants only and gender differences in reaction to facial resemblance may exist. In several previous studies, male, but not female, participants rated children's faces that resembled themselves as more attractive than children's faces that did not resemble themselves (Platek et al., 2002, 2003; Volk & Quinsey, 2002). In addition, facial resemblance affected hypothetical investment decisions, such as the willingness to adopt a child, more in men than in women (Platek et al., 2002, 2003; Volk & Quinsey, 2002, 2007). That males' judgements and decisions are more affected by facial resemblance than those of women may be caused by a greater need to rely on kinship cues such as facial resemblance to be certain of paternity. Contrary to women, who bear their child, throughout history, men could never be absolutely certain of paternity. Therefore, men may use their degree of facial resemblance with a child to make more accurate estimations of their likelihood of paternity, and, as a consequence, may adjust their behavior towards a child based on the degree of facial resemblance. As women do not need to rely on their degree of facial resemblance with a child to know whether or not it is their own, it makes sense that effects of facial resemblance would be more restricted in women than in men. Future studies, directly comparing females and males, could determine the extent of such gender differences, investigating for example whether gender differences extend to both small and large investment decisions and to both positive and negative judgments and behaviors.

Second, the design of our task may have contributed to differences between our and others' findings. Previous research has often used forced-choice tasks, in which participants were asked to indicate their preference for one of two or more simultaneously presented faces varying in facial resemblance (e.g. DeBruine, 2004a, 2004b; 2005; Krupp et al., 2008; Plateket al., 2002). In our study, we presented faces one at a time. Thus, participants evaluated each of the faces separately, and could therefore assign different faces similar ratings, decreasing differences between appraisal of non-resembling and resembling faces. Compared to ratings that allow for subtle variations in the appraisal of children's faces, forced choices may exaggerate effects of facial resemblance. In addition, we created morphs using the face of a nine- to eleven-year old child, whereas other studies have often used the faces of infants and young children (up to two years old) to create morphs (Bressan et al., 2009; DeBruine, 2004b; Platek et al., 2002, 2003, 2004; Platek, Kenaan, & Mohamed, 2005). As a consequence, the morphs in our study looked older than those used in previous research, which may have yield different results. Investment decisions are most likely to be made during infancy (Daly & Wilson, 1984) and infants require a large amount of care and parental contact, and therefore, facial resemblance in infants may evoke stronger reactions compared to facial resemblance in older children. In addition, parental investment and care are crucial for the survival of infants, but parental care becomes less crucial as children age and learn to be increasingly self-reliant and independent. This too might heighten the importance of facial resemblance in infants and young children In future research, morphs of infant and child faces of different ages could be included to examine whether and how the effects of facial resemblance are moderated by the age of the morphed face.

The present study has some limitations. First, we used a retrospective self-report questionnaire to assess participants' experiences with childhood love-withdrawal. Participants' memories might be biased by current experiences with their parents. Second, the 75%-resembling morph appeared slightly older than the 50%-resembling and the non-resembling morphs, which may have affected the results. However, participants usually rate faces of younger children higher in cuteness and attractiveness than faces of older children (Luo, Li, & Lee, 2011; Volk, Lukjanczuk, & Quinsey, 2007). Thus, our finding of a less negative appraisal of the 75%-resembling morph is unlikely to reflect an effect of perceived age. Third, it remains unclear to what extent facial resemblance created through morphing matches facial resemblance of actual genetic relatives. Genetically related faces are more similar than unrelated individuals' faces, but exactly what features individuals look at to estimate the degree of genetic relatedness has not yet been investigated systematically, and similarity in specific features (e.g. nose width) may be more important than overall similarity (DeBruine, Jones, Little, & Perrett, 2008; Maloney & Dal Martello, 2006). Moreover, we cannot exclude the possibility that some of the participants were aware of the study's hypothesis. In future research, it is important to explicitly ask the participants during debriefing about their ideas of the study hypothesis. Finally, we only included female participants in our study, due to concerns with regard to sample size and homogeneity, and to enable comparisons with planned future studies with mothers and their children.

Future research with larger samples should also include both males and females to examine possible gender differences in response to facial resemblance. In addition, future research should investigate the influence of phenotypic kinship cues on parental behavior to examine how differences in children's facial resemblance to their parents may affect real life behavior. Facial resemblance between fathers and their offspring may be positively related to fathers' investment of resources in their offspring (Alvergne, Faurie & Raymond, 2009; Prokop, Obertová, & Fedor 2010), and fathers' emotional closeness to their offspring (Alvergne et al., 2009). In addition, reactions to facial resemblance in expectant mothers' and fathers' may be examined, as hormonal changes in expectant mothers and fathers prior to the birth of a baby (Berg & Wynne-Edwards, 2001; Storey, Walsh, Quinton, & Wynne-Edwards, 2000; Rilling, 2013) may affect their reactions to facial resemblance in infants.

Finally, earlier research has shown interference between simultaneous processing of faces and names of familiar people (Young et al., 1986; Ferreira, Marful & Bajo, 2014). In future research, it could be investigated whether individuals with more love-withdrawal experiences differ from individuals with less experiences of love-withdrawal in the degree of interference between face perception and name retrieval. Moreover, future research could examine whether individuals with more love-withdrawal experiences do not only react stronger to faces of significant others, but also to other identifying information, including names.

Conclusion

In sum, we found evidence for an association between the neural processing of children's faces and appraisal of these faces. More extensive face processing was related to stronger reactions to facial resemblance (i.e. less negative appraisal with increasing facial resemblance) in participants who reported high levels of love-withdrawal. These results illustrate the long-term effects of early love-withdrawal. Thus, children of parents who regularly use love-withdrawal as a disciplinary strategy may become vigilant for information related to significant others, including their own offspring. Such enhanced vigilance may explain the stronger reactions to facial resemblance suggesting kinship.

References

Al-Gazali, L., Hamamy, H., & Al-Arrayad, S. (2006). Genetic disorders in the Arab world. *Bmj, 333*(7573), 831-834. doi: 10.1136/bmj.38982.704931.AE

Alvergne, A., Faurie, C., & Raymond, M. (2009). Father-offspring resemblance predicts paternal investment in humans. *Animal Behaviour, 78*(1), 61-69. doi: 10.1016/j.anbehav.2009.03.019

Alvergne, A., Faurie, C., & Raymond, M. (2010). Are parents' perceptions of offspring facial resemblance consistent with actual resemblance? Effects on parental investment. *Evolution and Human Behavior, 31*(1), 7-15. doi: 10.1016/j.evolhumbehav.2009.09.002

Anderson, K. G., Kaplan, H., & Lancaster, J. (1999). Paternal care by genetic fathers and stepfathers I: Reports from Albuquerque men. *Evolution and Human Behavior, 20*(6), 405-431. doi: 10.1016/S1090-5138(99)00023-9

Anderson, K. G., Kaplan, H., Lam, D., & Lancaster, J. (1999). Paternal care by genetic fathers and stepfathers II: Reports by Xhosa high school students. *Evolution and Human Behavior, 20*(6), 433-451. doi: 10.1016/S1090-5138(99)00022-7

Andrews, T. J., & Ewbank, M. P. (2004). Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *Neuroimage, 23*(3), 905-913. doi: 10.1016/j.neuroimage.2004.07.060

Andrews, T. J., & Schluppeck, D. (2004). Neural responses to Mooney images reveal a modular representation of faces in human visual cortex. *Neuroimage, 21*(1), 91-98. doi: 10.1016/j.neuroimage.2003.08.023

Assor, A., Roth, G., & Deci, E. L. (2004). The emotional costs of parents' conditional regard: A Self-Determination Theory analysis. *Journal of personality, 72*(1), 47-88. doi: 10.1111/j.0022-3506.2004.00256.x

Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., Riem, M. M., Tops, M., & Alink, L. R. (2012). Oxytocin decreases handgrip force in reaction to infant crying in females without harsh parenting experiences. *Social cognitive and affective neuroscience, 7*(8), 951-957. doi: 10.1093/scan/nsr067

Berg, S. J., & Wynne-Edwards, K. E. (2001, June). Changes in testosterone, cortisol, and estradiol levels in men becoming fathers. In *Mayo Clinic Proceedings* (Vol. 76, No. 6, pp. 582-592). Elsevier. doi: 10.4065/76.6.582

Beyers, W., & Goossens, L. (2003). Psychological separation and adjustment to university: Moderating effects of gender, age, and perceived parenting style. *Journal of adolescent research, 18*(4), 363-382. doi: 10.1177/0743558403018004003

Bressan, P., & Grassi, M. (2004). Parental resemblance in 1-year-olds and the Gaussian curve. *Evolution and Human Behavior, 25*(3), 133-141. doi: 10.1016/j.evolhumbehav.2004.03.001

Bressan, P., Bertamini, M., Nalli, A., & Zanutto, A. (2009). Men do not have a stronger preference than women for self-resemblant child faces. *Archives of sexual behavior, 38*(5), 657-664. doi: 10.1007/s10508-008-9350-0

Bowlby J. (1973/1985). “Attachment and loss”, in *Separation: Anxiety and Anger,* Vol. 2. Harmondsworth: Penguin Books.

Creighton, S. J., & Noyes, P. (1989). *Child Abuse Trends in England and Wales, 1983-1987*. London: National Society for the Prevention of Cruelty to Children.

Daly, M., & Wilson, M. (1980). Discriminative parental solicitude: A biological perspective. *Journal of Marriage and the Family*, 277-288.

Daly, M., & Wilson, M. (1984). A sociobiological analysis of human infanticide. *Infanticide: Comparative and evolutionary perspectives,* 487-502.

Daly, M., & Wilson, M. (1988a). Evolutionary social psychology and family homicide. *Science, 242*(4878), 519-524. doi: 10.1126/science.3175672.

Daly, M., & Wilson, M. I. (1988b). *Homicide.* Aldine, Hawthorne.

Daly, M., & Wilson, M. I. (1996). Violence against stepchildren. *Current Directions in Psychological Science, 5*(3), 77-80.

Daly, M., & Wilson, M. (2008). Is the” Cinderella Effect Controversial?. *A case study of evolution-minded research and critiques thereof,* 383-400.

DeBruine, L. M. (2002). Facial resemblance enhances trust. *The Royal Society, 269,* 1307–1312. doi: 10.1098/rspb.2002.2034

DeBruine, L. M. (2004a). Facial resemblance increases the attractiveness of same–sex faces more than other–sex faces. *Proceedings of the Royal Society of London B: Biological Sciences, 271*(1552), 2085-2090. doi: 10.1098/rspb.2004.2824

DeBruine, L. M. (2004b). Resemblance to self increases the appeal of child faces to both men and women. *Evolution and Human Behavior, 25*(3), 142-154. doi: 10.1016/j.evolhumbehav.2004.03.003

DeBruine, L. M. (2005). Trustworthy but not lust-worthy: Context-specific effects of facial resemblance. *Proceedings of the Royal Society of London B: Biological Sciences, 272*(1566), 919-922. doi: 10.1098/rspb.2004.3003

DeBruine, L. M., Jones, B. C., Little, A. C., & Perrett, D. I. (2008). Social perception of facial resemblance in humans. *Archives of Sexual Behavior, 37*(1), 64–77. doi: 10.1007/s10508-007-9266-0

Elliot, A. J., & Thrash, T. M. (2004). The intergenerational transmission of fear of failure. *Personality and Social Psychology Bulletin, 30*(8), 957-971.

Engell, A. D., & McCarthy, G. (2013). Probabilistic atlases for face and biological motion perception: an analysis of their reliability and overlap. *Neuroimage, 74,* 140-151. doi: 10.1016/j.neuroimage.2013.02.025

Evenhouse, E., & Reilly, S. (2000). *Is the wicked stepmother just a fairytale?.* Unpublished manuscript, Vanderbilt University Department of Economics, Nashville, TN.

Gelles, R. J., &Harrop, J. W. (1991). The risk of abusive violence among children with nongenetic caretakers. *Family Relations,* 78-83. doi: 10.2307/585662

Goldstein, M., & Heaven, P. C. (2000). Perceptions of the family, delinquency, and emotional adjustment among youth. *Personality and individual differences, 29*(6), 1169-1178. doi: 10.1016/s0191-8869 (99)00264-0

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature neuroscience, 7*(5), 555-562. doi: 10.1038/nn1224

Hamilton, W. D. (1963). The evolution of altruistic behavior. *The AmericanNaturalist, 97*(896), 354-356.

Hamilton, W. D. (1964). The genetical evolution of social behavior I, II. *Journal of Theoretical Biology, 7,* 17–52.

Hauber, M. E., & Sherman, P. W. (2001). Self-referent phenotype matching: theoretical considerations and empirical evidence. *Trends in neurosciences, 24*(10), 609-616. doi: 10.1016/S0166-2236(00)01916-0

Heckendorf, E., Huffmeijer, R., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2016). Neural Processing of Familiar and Unfamiliar Children's Faces: Effects of Experienced Love Withdrawal, but No Effects of Neutral and Threatening Priming. *Frontiers in Human Neuroscience, 10,* 231. doi: 10.3389/fnhum.2016.00231

Heckendorf, Bakermans-Kranenburg, Van IJzendoorn, & Huffmeijer (2018). *Neural responses to children's faces: Test-retest reliability of structural and functional MRI.* Submitted for publication.

Hepper, P. (2011). Kin recognition. *The Oxford handbook of evolutionary family psychology*, 211-229.

Hetherington, E. M., & Jodl, K. M. (1994). Stepfamilies as settings for child development. *Stepfamilies: Who benefits? Who does not*, 55-79.

Hoffman, M. L., & Saltzstein, H. D. (1967). Parent discipline and the child's moral development. *Journal of personality and social psychology*, *5*(1), 45.

Huffmeijer, R., Tops, M., Alink, L. R., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2011). Love-withdrawal is related to heightened processing of faces with emotional expressions and incongruent emotional feedback: evidence from ERPs. *Biological Psychology*, *86*(3), 307-313. doi: 10.1016/j.biopsycho.2011.01.003

Huffmeijer, R., Alink, L. R., Tops, M., Grewen, K. M., Light, K. C., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2013). The impact of oxytocin administration and maternal love withdrawal on event-related potential (ERP) responses to emotional faces with performance feedback. *Hormones and behavior, 63*(3), 399-410. doi: 10.1016/j.yhbeh.2012.11.008

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage, 17,* 825–841. doi: 10.1006/nimg.2002.1132

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *Neuroimage,* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09.015

Kaminski, G., Dridi, S., Graff, C., & Gentaz, E. (2009). Human ability to detect kinship in strangers' faces: effects of the degree of relatedness. *Proceedings of the Royal Society of London B: Biological Sciences, 276*(1670), 3193-3200. doi: 10.1098/rspb.2009.0677

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences, 361*(1476), 2109-2128. doi: 10.1098/rstb.2006.1934

Krupp, D. B., DeBruine, L. M., & Barclay, P. (2008). A cue of kinship promotes cooperation for the public good. *Evolution and Human Behavior, 29*(1), 49-55. doi: 10.1016/j.evolhumbehav.2007.08.002

Lehmann, L., & Perrin, N. (2002). Altruism, dispersal, and phenotype-matching kin recognition. *The American Naturalist, 159*(5), 451-468. doi: 10.1086/339458

Lieberman, D., Tooby, J., & Cosmides, L. (2007). The architecture of human kin detection. *Nature, 445*, 727–731.doi: 10.1038/nature05510

Locke, L. M., & Prinz, R. J. (2002). Measurement of parental discipline and nurturance. *Clinical psychology review, 22*(6), 895-929. doi: 10.1016/S0272-7358(02)00133-2

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. doi: 10.1016/j.jecp.2011.04.002

Malkin, C. M., Lamb, M. E. (1994). Child maltreatment: test of sociobiological theory. *Journal of Comparative Family Studies, 25,*121-134.

Maloney, L. T., & Dal Martello, M. F. (2006). Kin recognition and the perceived facial similarity of children. *Journal of Vision, 6*(10), 4-4. doi: 10.1167/6.10.4

Marlowe, F. (1999). Male care and mating effort among Hadza foragers. *Behavioral Ecology and Sociobiology, 46*(1), 57-64. doi: 10.1007/s002650050592

Patrick, R. B., & Gibbs, J. C. (2007). Parental expression of disappointment: should it be a factor in Hoffman's model of parental discipline?. *The Journal of genetic psychology, 168*(2), 131-146. doi: 10.3200/gntp.168.2.131-146

Platek, S. M., Burch, R. L., Panyavin, I. S., Wasserman, B. H., & Gallup, G. G. (2002). Reactions to children's faces: Resemblance affects males more than females. *Evolution and Human Behavior, 23*(3), 159-166. doi: 10.1016/S1090-5138(01)00094-0

Platek, S. M., Critton, S. R., Burch, R. L., Frederick, D. A., Myers, T., & Gallup, G. G., Jr. (2003). How much paternal resemblance is enough? Sex differences in the reaction to resemblance but no in ability to detect resemblance. *Evolution andHuman Behavior, 24*, 81–87. doi: 10.1016/S1090-5138(02)00117-4

Platek, S. M., Keenan, J. P., & Mohamed, F. B. (2005). Sex differences in the neural correlates of child facial resemblance: An event-related fMRI study. *Neuroimage, 25*, 1336–1344. doi: 10.1016/j.neuroimage.2004.12.037

Platek, S. M., Raines, D. M., Gallup, G. G., Mohamed, F. B., Thomson, J. W., Myers, T. E., ... & Arigo, D. R. (2004). Reactions to children's faces: Males are more affected by resemblance than females are, and so are their brains. *Evolution and Human behavior, 25*(6), 394-405. doi: 10.1016/j.evolhumbehav.2004.08.007

Porter, R. H., & Moore, J. D. (1981). Human kin recognition by olfactory cues. *Physiology &Behavior, 27*(3), 493-495.

Prokop, P., Obertová, Z., & Fedor, P. (2010). Paternity cues and mating opportunities: what makes fathers good?. *Actaethologica, 13*(2), 101-107. doi: 10.1007/s10211-010-0079-0

Renk, K., McKinney, C., Klein, J., & Oliveros, A. (2006). Childhood discipline, perceptions of parents, and current functioning in female college students. *Journal of adolescence, 29*(1), 73-88. doi: 10.1016/j.adolescence.2005.01.006

Riem, M. M., van IJzendoorn, M. H., Tops, M., Boksem, M. A., Rombouts, S. A., & Bakermans-Kranenburg, M. J. (2013). Oxytocin effects on complex brain networks are moderated by experiences of maternal love withdrawal. *EuropeanNeuropsychopharmacology, 23*(10), 1288-1295. doi: 10.1016/j.euroneuro.2013.01.011

Rilling, J. K. (2013). The neural and hormonal bases of human parentalcare. *Neuropsychologia, 51*(4), 731-747. doi: 10.1016/j.neuropsychologia.2012.12.017

Saha, N., Hamad, R. E., & Mohamed, S. (1990). Inbreeding effects on reproductive outcome in a Sudanese population. *Human heredity, 40*(4), 208-212. doi: 10.1159/000153932

Schludermann, E. H., & Schludermann, S. M. (1988). Children's report on parent behavior (CRPBI-108, CRPBI-30) for older children and adolescents (Tech. Rep.). *Winnipeg, MB, Canada: University of Manitoba, Department of Psychology*.

Schludermann, S. M., & Schludermann, E. H. (1983). Sociocultural change and adolescents' perceptions of parent behavior. *Developmental Psychology, 19*(5), 674.

Schmeeckle, M. (2007). Gender dynamics in stepfamilies: Adult stepchildren's views. *Journal of Marriage and Family, 69*(1), 174-189. doi: 10.1111/j.1741-3737.2006.00352.x

Smith S. M., Jenkinson M., Woolrich M. W., Beckmann C. F., Behrens T. E., Johansen-Berg H., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage,* 23, S208–219. doi: 10.1016/j.neuroimage.2004.07.051

Storey, A. E., Walsh, C. J., Quinton, R. L., & Wynne-Edwards, K. E. (2000). Hormonal correlates of paternal responsiveness in new and expectant fathers. *Evolution and Human Behavior, 21*(2), 79-95. doi: 10.1016/S1090-5138(99)00042-2

Tabachnick, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics,* 4th Edn. Boston: Allyn and Bacon.

Temrin, H., Buchmayer, S., & Enquist, M. (2000). Step–parents and infanticide: new data contradict evolutionary predictions. *Proceedings of the Royal Society of London B: Biological Sciences, 267*(1446), 943-945. doi: 10.1098/rspb.2000.1094

Tifferet, S., Jorev, S., & Nasanovitz, R. (2010). Lower parental investment in stepchildren: The case of the Israeli" Great Journey.". *Journal of Social, Evolutionary, and Cultural Psychology, 4*(2), 62. doi: 10.1037/h0099300

Van IJzendoorn, M. H., Huffmeijer, R., Alink, L. R., Bakermans-Kranenburg, M. J., & Tops, M. (2011). The impact of oxytocin administration on charitable donating is moderated by experiences of parental love-withdrawal. *Frontiers in psychology, 2,* 258. doi: 10.3389/fpsyg.2011.00258

Van IJzendoorn, M.H., Euser, E.M., Prinzie, P., Juffer, F., & Bakermans-Kranenburg, M.J. (2009). Elevated risk of child maltreatment in families with stepparents but not with adoptive parents. *Child Maltreatment, 14,* 369-375. doi: 10.1177/1077559509342125

Volk, A. A., Lukjanczuk, J. L., & Quinsey, V. L. (2007). Perceptions of child facial cues as a function of child age. *Evolutionary Psychology, 5*(4), doi: 147470490700500409.

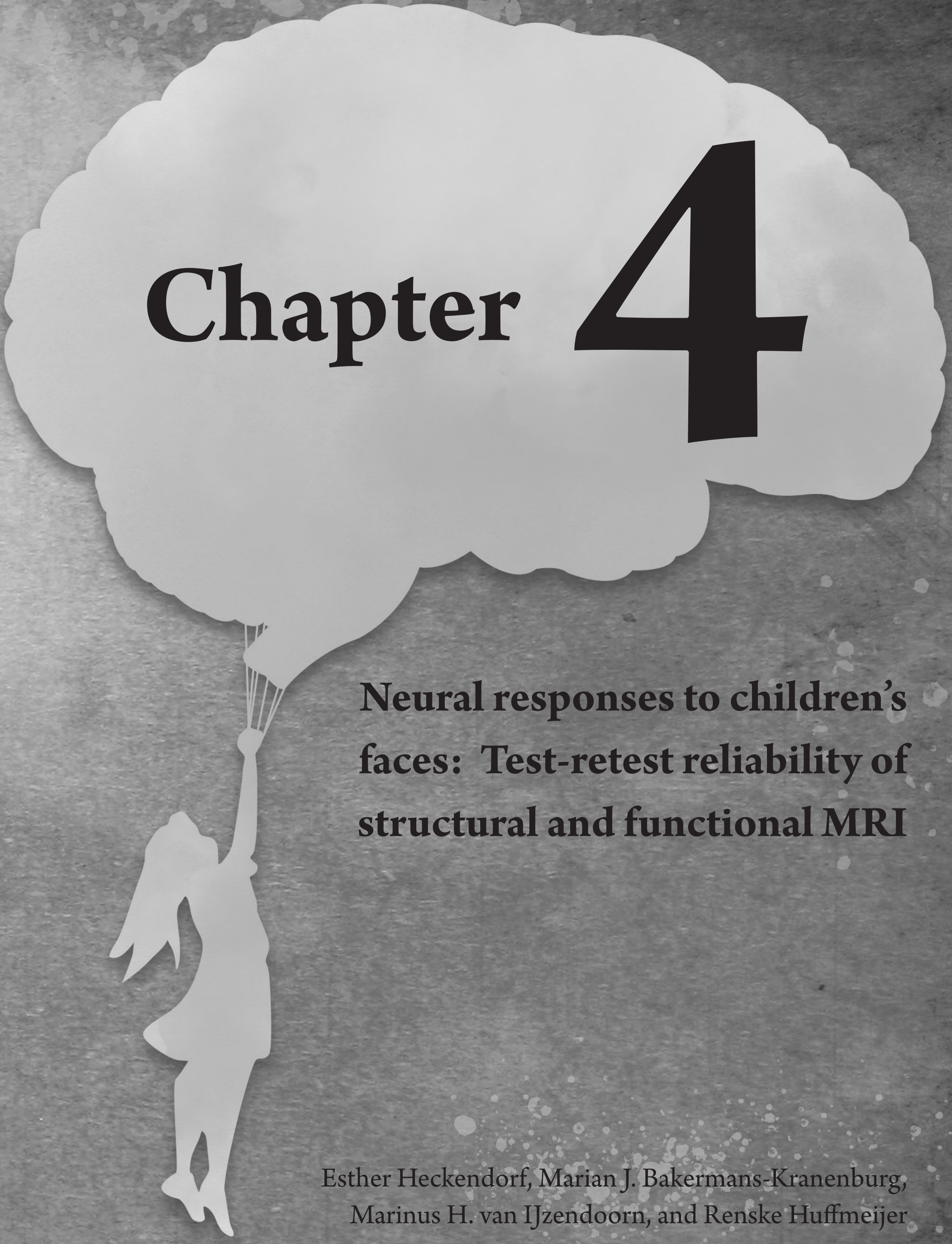
Volk, A., & Quinsey, V. L. (2002). The influence of infant facial cues on adoption preferences. *Human Nature, 13*(4), 437-455. doi: 10.1007/s12110-002-1002-9.

Volk, A. A., & Quinsey, V. L. (2007). Parental investment and resemblance: Replications, refinements, and revisions. *Evolutionary Psychology, 5*(1), doi: 147470490700500101.

Young, A. W., McWeeny, K. H., Ellis, A. W., & Hay, D. C. (1986). Naming and categorizing faces and written names. *The Quarterly Journal of Experimental Psychology Section A, 38*(2), 297-318.

Yovel, G., & Kanwisher, N. (2005). The neural basis of the behavioral face-inversion effect. *Current Biology, 15*(24), 2256-2262. doi: 10.1016/j.cub.2005.10.072

Zvoch, K. (1999). Family type and investment in education: A comparison of genetic and stepparent families. *Evolution and Human Behavior, 20*(6), 453-464. doi: 10.1016/S1090-5138(99)00024-0.



Chapter 4

Neural responses to children’s faces: Test-retest reliability of structural and functional MRI

Esther Heckendorf, Marian J. Bakermans-Kranenburg,
Marinus H. van IJzendoorn, and Renske Huffmeijer

Manuscript submitted for publication

Abstract

Functional MRI (fMRI) is commonly used to investigate the neural mechanisms underlying psychological processes and behavioral responses. However, to draw well-founded conclusions from fMRI studies, more research on the reliability of fMRI is needed. For this purpose, we invited a sample of 41 female students to participate in two identical fMRI-sessions, separated by 5 weeks on average. To investigate the potential effect of left-handedness on the stability of neural activity, we oversampled left-handed participants (N=20). Inside the scanner, we presented photographs of familiar and unfamiliar children's faces preceded by neutral and threatening primes to the participants. We calculated intra-class correlations (ICCs) to investigate the test-retest reliability of peak activity in areas that showed significant activity during the first session (primary visual cortex, fusiform face area, inferior frontal gyrus, and superior temporal gyrus). In addition, we examined how many trials were needed to reliably measure the effects. Across all participants, only fusiform face area activity in response to faces showed good test-retest reliability (ICC=.71). All other test-retest reliabilities were low (.01 ≤ ICC ≤ .35). Reliabilities varied only slightly with increasing numbers of trials, with no consistent increase in ICCs. Test-retest reliabilities for left-handed participants (.28 ≤ ICC ≤ .66) were generally somewhat higher than for right-handed participants (-.13 ≤ ICC ≤ .75), but not statistically significant. Our study thus shows good test-retest reliability for FFA activity in response to faces, but low test-retest reliability for other contrasts and areas.

Introduction

Twenty-five years after the first functional MRI (fMRI) experiment was conducted (Belliveau et al., 1991), fMRI has grown into a universally used method to study the neural correlates of both psychological and behavioral responses to visual or auditory stimuli. However, firm conclusions can only be drawn from fMRI experiments if the measurements are valid, that is, assess what they are supposed to measure, and test-retest reliable, that is, provide stable results over time. Reliability is usually considered a prerequisite for validity (Feldt & Brennan, 1989; Gay, 1987).

In earlier studies, fMRI test-retest reliability was investigated using various tasks and experimental designs (see Bennett & Miller, 2010). Bennett and Miller (2010) computed an average intraclass correlation coefficient (ICC) of .50 across 13 earlier fMRI reliability studies, but report substantial variation across studies, with ICCs ranging from .16 to .88. This considerable variance in ICC values for task-related fMRI measures may be caused by technical factors (e.g. magnet strength of the scanner), the process under investigation (e.g. visual processing, memory) task design (e.g. block design vs. event-related design), sample characteristics, and the time interval between the two assessments (Bennett & Miller, 2010). In the current study, we investigated the influence of some of these factors on test-retest reliability in a face processing paradigm.

A Web of Science search with the search terms 'face' and 'fMRI' (WoS, September 1st, 2016) results in more than 4,000 hits for studies conducted during the last 20 years, which illustrates how common the investigation of face processing in neuroimaging research is. Nevertheless, studies assessing test-retest reliability for face processing tasks are surprisingly rare. The existing studies of test-retest reliability of fMRI activity in face processing paradigms focused on the processing of faces with emotional expressions. In most of these studies, poor test-retest reliability of amygdala activity was reported (Lipp, Murphy, Wise, & Caseras, 2014; Plichta et al., 2012; Sauder, Hajcak, Angstadt, & Phan, 2013; Van den Bulk et al., 2013), with the exception of two studies reporting fair to excellent amygdala test-retest reliability (Cao et al., 2014; Schacher et al., 2006). Test-retest reliability for regions other than the amygdala revealed fair reliability for prefrontal cortex activity (Van den Bulk et al., 2013), and fair to good test-retest reliability for fusiform face area (FFA) activity (Sauder et al., 2013). The sample sizes of these studies were mostly small, ranging from 12 to 27 participants. In fact, neuroscientific studies tend to be underpowered in general (due to small sample sizes and/or small effects; Button et al., 2013), and fMRI reliability studies are no exception. In the review of Bennett and Miller (2010), the overall sample size across 63 studies was 11, with many studies using fewer than 10 subjects for reliability measures. fMRI reliability studies with larger sample sizes are thus badly needed. Here, we aimed to fill this gap by conducting a reliability study with a larger sample (N=41).

Moreover, to the best of our knowledge, no studies have yet investigated the test-retest reliability of a face processing paradigm with faces with neutral expressions only, although these are regularly used in fMRI-research. Here, we specifically address the test-retest reliability of fMRI activity during a face processing task with faces without emotional expressions (Heckendorf et al., 2016). We examine the test-retest reliability of fMRI data acquired during two sessions separated by a period of 4-12 weeks. Because the stability of significant activity is particularly informative in light of the reproducibility of neuroimaging research, we computed between-session reliability of effects that were significant in session 1 (for details see Heckendorf, Bakermans-Kranenburg, Van IJzendoorn, & Huffmeijer, 2016). We targeted the following regions of interest (ROIs): inferior frontal gyrus (IFG), superior temporal gyrus (STG), fusiform face area (FFA), and primary visual cortex (V1). However, as limited reliability within a single session may negatively affect test-retest reliability, we also computed within-session reliability for both session 1 and session 2. Based on the meta-analysis of Bennett and Miller (2010), we expected fair test-retest reliability values for the fMRI data in our study.

We examined effects of two specific factors on reliability. First, we examined whether test-retest reliability differs between left- and right-handed participants. Left-handers are frequently excluded from neuroimaging studies to prevent the introduction of unwanted noise in group statistics, that would, for instance, be caused by potential differences in lateralization between left- and right-handers (Willems, Van der Haegen, Fisher, & Francks, 2014). However, about 10% of humans are left-handed, and thus left-handers represent a significant proportion of the human population (McManus, 2009). Thus, we aim to examine to what extent brain activity of left-handed participants can be measured as reliably as right-handed participants' brain activity. Second, we examined the influence of task length, and thus the number of volumes scanned per participant. In resting-state fMRI, both increasing the number of volumes and increasing the time over which these volumes are acquired have been shown to improve within- and between-session reliabilities (Birn et al., 2013). Likewise, in ERP-studies, the reliability of averaged ERPs can benefit from increasing the numbers of trials (Huffmeijer, Bakermans-Kranenburg, Alink, & Van IJzendoorn, 2014). Whether increasing the number of trials of a task significantly improves test-retest reliability of task fMRI-data has not yet been studied systematically.

Finally, we assessed test-retest reliability for several measures of structural MRI as a comparison to fMRI. We focused on measures of gray and white matter volume as well as volumetric measures of two subcortical structures: the amygdala and the thalamus. We expected good to excellent reliability of all volumetric measures, in accordance with earlier research (Bartzokis et al., 1993; Convit et al., 1999; Morey et al., 2010).

Materials and methods

Participants.

We invited 49 female undergraduate and graduate students with an average age of 21.73 years (*SD* = 2.55, range 18-28 years) for two experimental sessions, 4-12 weeks (*M* = 4.61, *SD* = 1.68 weeks) apart. Exclusion criteria were MRI contraindications, pregnancy, current psychiatric and neurological disorders, severe head injury, current alcohol or drug abuse, and chronic use of medication (except contraceptives). Data of two of the participants could not be included in test-retest reliability calculations, because they only completed the first session of the experiment. In addition, data of six participants were excluded from analyses because of excessive head movements (> 3mm; *n*= 2) or falling asleep during the fMRI recording (*n*= 4). Thus, our final sample consisted of 41 participants aged 21.81 years on average (*SD* = 2.67; range 18-28 years). The Ethics Committee of the Leiden University Medical Center approved the study and all participants signed informed consent at the beginning of the first session. Participation was rewarded with 40 €. No anomalies were found on the structural MRI scans of the participants.

Procedure

Prior to the first session, participants' completed Van Strien's (1992) 10-item Handedness Questionnaire, which measures hand preference during execution of several tasks (e.g., "Which hand do you use to brush your teeth?"). Items are scored on a 3-point scale (left hand, both hands, right hand) ranging from -1 to 1. Total scores can thus vary between -10 and +10. Based on their scores, we divided the participants into two groups: participants with a score of +1 or higher were defined as right-handed (N= 21), and participants with a score of -1 or lower were classified as left-handed (N= 20). We oversampled left-handed participants to investigate the potential effect of left-handedness on the stability of neural activity.

We asked participants to abstain from alcohol and excessive physical activity during the last 24 hours and from caffeine during the last 12 hours before the start of each session. In session 1, participants filled out the Children's Report of Parental Behavior Inventory (CRPBI-30, Schludermann & Schludermann, 1983; Beyers & Goossens, 2003) and the Interpersonal Reactivity Index (Davis, 1980; De Corte et al., 2007). Results relating to these questionnaires and fMRI data obtained during the first session have been reported elsewhere (Heckendorf et al., 2016). At the beginning of each session, the MRI procedure was explained to the participants. Inside the scanner, foam inserts were placed between the head coil and the participant's head to minimize head movements. Within the scanner, participants completed a priming task (see below), during which visual stimuli were projected onto a screen placed outside the opening of the scanner bore. Participants viewed the screen through a mirror fixed to the head coil. At the end of the second session, participants completed a task in which they judged several characteristics of various faces (data to be reported

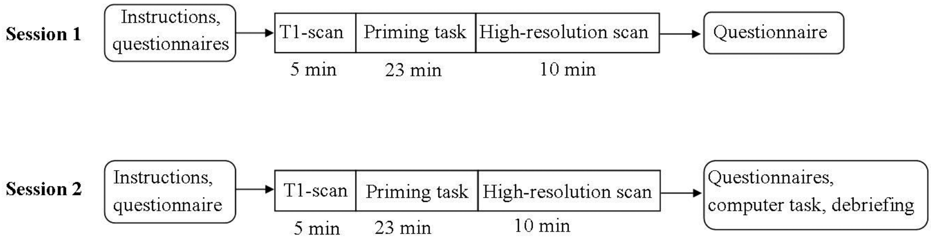


Figure 1. Schematic overview of session 1 and session 2 and the scan procedures. Scan procedures were identical for the two sessions.

elsewhere). Subsequently, participants were debriefed about the nature of the priming task. Figure 1 shows a schematic overview of the procedures in each session.

Experimental task

Inside the scanner, subjects completed a priming task consisting of 234 trials. The priming task was set-up in an event-related design. All stimuli were shown in the center of the screen on a black background. On all trials, a colored, circular pattern was used for forward and backward masking of the primes to prevent conscious perception of the primes. The mask was matched for size and average luminosity of the primes. During each trial, a fixation cross was presented (1,800-10,600 ms), followed by the mask (presented for 484 ms), a neutral or a threatening prime (presented for 16 ms), again the mask (presented for 100 ms) and an unfamiliar-looking, a familiar-looking or a scrambled face (presented for 2,000 ms). Thus, the priming task consisted of six conditions: a familiar-looking face presented after a neutral prime (neutral-familiar), a familiar-looking face presented after a threatening prime (threat-familiar), an unfamiliar-looking face presented after a neutral prime (neutral-unfamiliar), an unfamiliar-looking face presented after a threatening prime (threat-unfamiliar), a scrambled face presented after a neutral prime (neutral-scrambled), and a scrambled face presented after a threatening prime (threat-scrambled). We presented stimulus sequences (mask-prime-mask-[scrambled]face) in quasi-random order with the following restrictions: the same prime could not be presented more than twice in a row, the same face could not be presented more than four times in a row, and the same condition could not be presented more than two times in a row. In total, the priming task consisted of 13 neutral and 13 threatening primes, that were each presented 3 times with each face, resulting in 39 (3*13) trials per condition. Participants had to press a button every 11-13 trials to continue the task to verify that they remained alert. On average, the task took 23 minutes. Figure 2 illustrates a trial of the priming task.

Primes

In previous research, Nummenmaa, Hirvonen, Parkkola, & Hetanen (2008) developed pairs of neutral and threatening photographs that were matched on luminosity, global energy, contrast density, and complexity, and depicted two persons in comparable proximity to each other. Threatening scenes portrayed interpersonal attack scenes (e.g. one person strangling the other), whereas neutral scenes depicted emotionally neutral situations (e.g. two persons having a conversation). We used these neutral and threatening photographs as primes for our study with the objective to investigate subliminal processing of neutral and threatening stimuli. For a detailed description concerning the selection of the neutral and threatening pairs and the visibility of the primes in our study see Heckendorf et al. (2016).

Facial stimuli

We morphed a photograph of a child's face (unfamiliar to the participant) with (i) a photograph of an unknown female's face and (ii) a photograph of the participant's own face to create unfamiliar- and familiar-looking children's

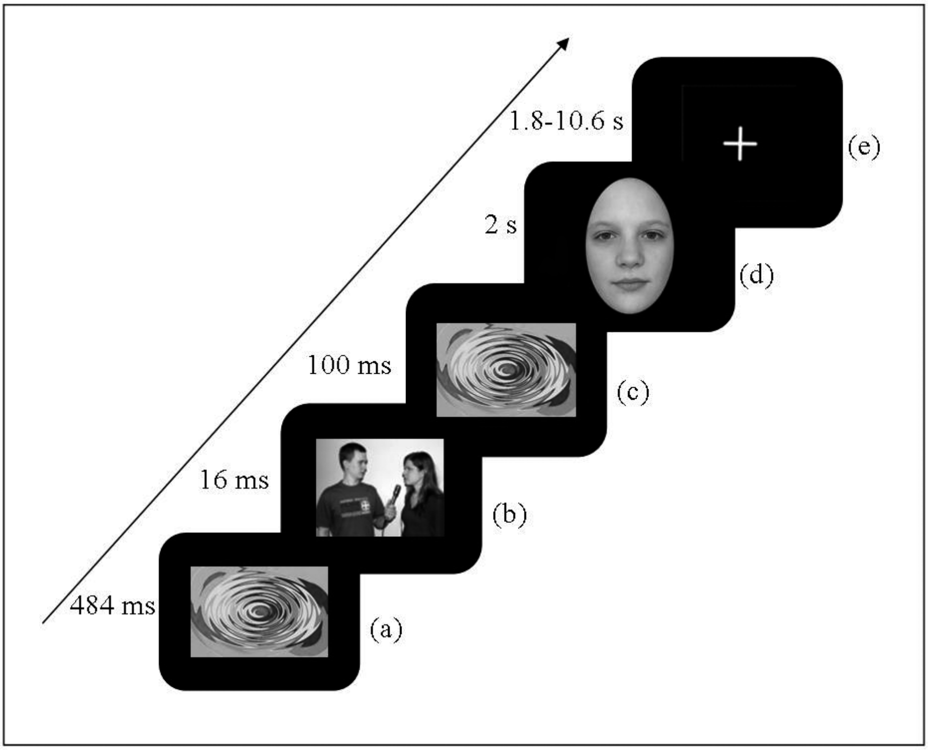


Figure 2. The priming task. A neutral or threatening prime (b) was presented for 16 ms on the screen, concealed by a mask presented immediately before (a) and after (c) the prime. The second mask was followed by an unfamiliar-looking, a familiar-looking or a scrambled face (d). During intertrial intervals, a fixation (e) cross was presented.

faces. For this purpose, we asked participants prior to the first session to provide a full-color digital photograph of themselves that met the following criteria: picture on a light and uniform background, showing their face (full frontal) and neck only, with a neutral facial expression, and no piercings, make-up or glasses. To create the unfamiliar-looking morphs, we used two full color, full frontal photographs of two female Caucasian faces (aged 24 and 25 years) with neutral facial expression, no jewelry or glasses and unfamiliar to the participant. For half of the participants, we created unfamiliar-looking morphs with female face 1 for session 1 and female face 2 for session 2, and for the other half vice versa. Full color, full frontal photographs of six 9-11 year old children (three boys and three girls, all Caucasian [but slightly varying in skin color], all unfamiliar to the participants, with neutral facial expression, no jewelry or glasses) were available for morphing. We used a picture of a female child to create morphs for half of the participants and a picture of a male child to create morphs for the other half of the participants. Within genders, the child that best matched the participant's skin color and face-shape was selected for ease of morphing. We used the photograph of the same child to create unfamiliar-looking and familiar-looking morphs for a participant. One familiar-looking and two-unfamiliar-looking morphs were created for the two sessions, because using the same unfamiliar-morph for both session would have increased familiarity with the unfamiliar-looking face in session 2 compared to session 1. Using a different unfamiliar-looking face in session 2 ensured that participants' familiarity with the unfamiliar-looking face was kept constant across sessions, in order to avoid effects on test-retest reliability. To generate the morphs, all photographs were first resized to 448 x 560 pixels and edited using Adobe Photoshop CS: External features (i.e. hair and ears) were removed and the pictures were pasted on a black background. Next, Fantamorph 5 Deluxe was used to create the morphs. We created familiar-looking morphs that consisted for 50% of the participant's face and for 50% of an unknown child's face, and unfamiliar-looking morphs that consisted for 50% of the unknown female's face and for 50% of the child's face. The resulting morphs looked somewhat older than the 9-11 year olds used for morphing, and appeared to be about 14 years old (see Heckendorf et al., 2016). Finally, a scrambled face was created for each participant from the familiar-looking morph by randomly rearranging blocks of 9 x 9 pixels using Matlab R2012B.

Image acquisition

Images were acquired at the Leiden University Medical Center on a 3-T Philips Achieva MRI system (Philips Medical Systems, Best, Netherlands) with a 32-channel SENSE (Sensitivity Encoding) head coil. An event-related design with 680 T2*-weighted whole-brain echo planar images (EPI, repetition time (TR) = 2200ms, echo time (TE) = 30ms, flip angle = 80°, 38 transverse slices, descending acquisition order, voxelsize = 2.75 x 2.75 x 3.025 mm^3 with a 10% interslice gap, field of view (FOV) = 220 x 114.675 x 220 mm3) was used for the functional scans. To avoid magnetic saturation effects, the first four functional scans were discarded. In addition, an anatomical 3D T1-weighted scan (TR = 9.825ms, TE = 4.605ms, inversion time (TI) = 1050 ms, shot interval = 1932 ms, flip angle = 8°, 140 transverse slices, voxelsize 0.875 x 0.875 x 1.2 mm^3, FOV = 224 x 168 x 177.333 mm^3), and a high-resolution T2*-weighted EPI-image (TR = 2200ms, TE = 30ms, flip angle = 80°, 84 transverse slices, voxel size = 1.964 x 1.964 x 2 mm^3, FOV = 220 x 168 x 220 mm^3) were obtained during each session for coregistration

fMRI Data Analysis

Data-analyses were carried out using FSL (FMRIB's Software Library1) FEAT (FMRI Expert Analysis Tool) version 5.0.4, part of Jenkinson, Beckmann, Behrens, Woolrich, & Smith (2012) and Smith et al. (2004). Data obtained during sessions one and two were processed identically. Because we were interested in potential effects of task length on the reliability of MRI data, and in within-session reliability, analyses were performed on both the complete datasets and several subsets of data: 1) data obtained during the first third (78 trials) and 2) first two thirds (156 trials) of the task, as well as data collected during 3) the second third and 4) final third of the task only. We made use of the Fslroi toolbox of FSL (FMRIB's Software Library) to create the different subsets of the data. Subsequently, data of the different subsets were processed identically to the data of the whole task.

Four pre-statistics processing steps were applied to the data: motion correction (MCFLIRT; Jenkinson, Bannister, Brady, & Smith 2002), non-brain removal (using BET; Smith, 2002), spatial smoothing using a Gaussian kernel with a full-width-at-half-maximum of 6 mm, and high-pass temporal filtering with a high-pass filter cutoff of 100 s. Subsequently, functional images were registered to the high-resolution EPI-image, which was then registered to the

3D T1-weighted scan, and then to the 2 mm isotropic MNI-152 standard space image (T1 standard brain averaged over 152 subjects; Montreal Neurological Institute, Montreal, QC, Canada; Jenkinson et al., 2002). Functional images of session 1 were registered to the high-resolution EPI-image and the 3D T1-weighted scan of session 1. Functional images of session 2 were registered to the high-resolution EPI-image and the 3D T1-weighted scan of session 2. General linear model analyses in native space were performed to examine functional activity in response to the stimuli. Because primes and masks were displayed on the screen for very short durations and time-locked to the presentation of the faces, hemodynamic responses to the individual stimuli within a mask-prime-mask-face sequence overlapped substantially and summed to a total, summed hemodynamic response to the stimulus sequence. Hence, we treated the presentation of a mask-prime-mask-face sequence as a single stimulation period. Thus, we modeled the different conditions (threat-familiar, threat-unfamiliar, threat-scrambled, neutral-familiar, neutral-unfamiliar, and neutral-scrambled) and participants' button press responses as seven explanatory variables using the Custom (3 column format) wave function convolved with a double gamma hemodynamic response function. The temporal derivatives of the explanatory variables were included in the model, yielding 14 regressors.

As described in Heckendorf et al. (2016), ROI- and whole-brain analyses of session 1 revealed greater activity in the primary visual cortex (V1) in all conditions of the priming task (threat-familiar, threat-unfamiliar, threat-scrambled, neutral-familiar, neutral-unfamiliar, neutral-scrambled) compared to fixation cross. Additionally, compared to unfamiliar faces, familiar faces evoked enhanced activity in the right IFG and in bilateral FFA, and unfamiliar faces, compared to familiar faces, elicited increased activity in bilateral STG. ICC values can be affected by systematic differences in brain activity between the sessions. Thus, to identify a possible session effect, we conducted separate ROI- and whole-brain analyses in which we compared activity in session 1 with activity in session 2 (for this purpose, we added a comparison of the two sessions to the model described in Heckendorf et al., 2016). The whole-brain and ROI-analyses did not reveal any significant session effects.

To analyze test-retest reliability of activity within the brain areas showing significant effects in session 1, we created two types of ROI-masks: a mask matching the area showing significant activity (differences) in session 1 (functional mask) and an anatomical mask. Anatomical masks for the IFG and

STG were defined using the Harvard-Oxford Cortical Structures Atlas. For V1, an anatomical mask was defined using the Juelich Histological Atlas (both Atlases are implemented in FSL version 5.0.4). Three binarized, anatomical masks consisting of voxels belonging to V1, left or right IFG and STG respectively with a probability of at least 25% were created in 2 mm isotropic MNI-152 standard space (Jenkinson et al., 2002). As the FFA is an area within the fusiform gyrus defined by its preferential responding to faces, we created an 'anatomical' mask of the FFA using the probability map obtained for a localizer task in an earlier study (N=124) for the contrast faces vs. scenes (Engell & McCarthy, 2013). Subsequently, we binarized and thresholded this contrast image (only voxels with a 25% probability to be significantly activated in the faces vs. scenes contrast included) in 2 mm isotropic MNI-152 standard space (Jenkinson et al., 2002). Next, we defined an anatomical mask of the fusiform gyrus using the Harvard-Oxford Cortical Structures Atlas. A binarized, anatomical mask including only voxels belonging to the right or left fusiform gyrus with a probability of at least 25% was created in 2 mm isotropic MNI-152 standard space (Jenkinson et al., 2002). Finally, we multiplied this mask with the thresholded face vs. scene contrast image to obtain an 'anatomical' mask for the FFA.

To create functional masks, we binarized and thresholded ($Z > 2.3$) the contrast images of significant effects obtained in session 1, using Fslstats (FMRIB's Software Library). Next, we multiplied these thresholded contrast images with the anatomical masks that we had created before (as significant clusters of activity sometimes extended over several anatomical areas or, conversely, covered only a part of the anatomical region). This resulted in the following four functional masks: 1) IFG and 2) STG masks matching activity obtained for the contrast familiar vs. unfamiliar, 3) FFA mask matching the combined activity obtained for the contrasts familiar vs. unfamiliar and face vs. scrambled, and 4) V1 activity obtained for the combined activity of the different conditions of the priming task vs. fixation cross. Subsequently, we used Featquery (Smith et al., 2004) to extract participants' mean, median and maximum cope (contrast of parameter estimates) values for each ROI and contrast of interest. These values were exported to IBM SPSS Statistics 23 for further analysis. As the reliability of significant activity was our main interest, results for functional masks are presented in the results section and results for anatomical masks can be found in the Supplementary material.

Structural MRI analyses

The anatomical 3D T1-weighted scans of session 1 and session 2 were segmented into gray matter, white matter and cerebrospinal fluid using FSL (FMRIB's Software Library) FAST (fMRI Automated Segmentation Tool) version 5.0.4, part of Jenkinson et al. (2012) and Smith et al. (2004). We used the Fslstats toolbox of FSL (FMRIB's Software Library) to extract measures of gray and white matter volume (in ml) from the partial volume maps that were created with FAST. We segmented the left and right thalamus and amygdala using FIRST (FMRIB's integrated registration and segmentation tool) and extracted measures of tissue volume (ml) of the left thalamus, right thalamus, left amygdala, and right amygdala using Fslstats. All values were exported to IBM SPSS Statistics 23 for further analysis. For one participant who was included in the fMRI test-retest reliability analyses, only the 3D T1-weighted scan of session 2 was available (due to a large artifact on the scan from session 1), and thus, it was not possible to calculate test-retest reliability of the structural MRI data for this participant. For this participant, the 3D T1-scan of session 2 was used for the registration of the functional images of both session 1 and session 2.

Intraclass correlation

To investigate test-retest reliability we calculated ICCs (2-way mixed model, single measures, absolute agreement) between values (volumetric measures for structural MRI and copes for fMRI data) obtained during sessions 1 and 2. According to Fleiss (1986) ICC values below .40 indicate poor reliability, values between .40 and .58 fair reliability, values between .59 and .75 good reliability and values above .75 indicate excellent reliability.

We calculated ICCs for the following values: 1) gray and white matter volumes, and tissue volumes of left thalamus, right thalamus, left amygdala, and right amygdala obtained from structural MRI data, and 2) mean, median and maximum values within the IFG, FFA and STG for the contrast familiar vs. unfamiliar, within the FFA for the contrast face vs. scrambled, and within V1 for the different conditions of the priming task vs. fixation cross. In general, ICCs were largest for maximum values from functional ROIs. Therefore, we present only ICCs for the maximum values in the results section (results for mean and median values are presented in the supplementary material), and focus on ICCs for maximum values in analyses of potential effects of task length and handedness. To investigate potential effects of the number of trials included, we calculated ICC values for different subsets of the data. For this purpose, we compared ICC values of the first third (78 trials), the first two thirds of the task (156 trials), and the complete task (234 trials). Finally, to investigate potential effects of handedness on test-retest reliability, we computed ICCs (complete task) for left- and right handed participants separately. We computed Fisher's r to z transformation to examine potential effects of handedness. To control for multiple testing, we applied the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995). For the structural data, we controlled for the number of tests conducted to examine test-retest reliability of volumetric measures for structural MRI. For functional data, we applied the Benjamini-Hochberg procedure separately for each of the functional processes that we investigated (visual processing [stimuli vs. fixation cross in V1], processing of familiarity [contrast: familiar vs. unfamiliar in IFG, STG, and FFA], and face processing [contrast: face vs. scrambled in FFA]).

In addition to ICCs for absolute agreements, we also computed ICCs for consistency (2-way mixed model, single measures, consistency) for maximum values from functional ROIs, to ensure that our ICCs for absolute agreements were not negatively affected by systematic activity differences between the sessions. Results related to ICCs obtained for consistency are presented in the Supplementary material (Table S4.5). We also investigated within-session reliability. For this purpose, within-session reliability was calculated as the ICC across values obtained during the first third, second third, and final third of the task in session 1 (each subset consisting of 78 trials). Likewise, within-session reliability of session 2 was calculated as the ICC across values obtained during the first third, second third, and final third of the task in session 2. Results regarding within-session reliability are presented in the Supplementary material. Because deviations from normal distributions and outliers may influence the ICCs, we examined skewness and kurtosis and possible outliers ($Z \geq 3.29$) for all distributions prior to test-retest reliability calculations. For distributions with outliers, we calculated test-retest reliability both with and without outliers. Removing outliers from the sample did not lead to substantial changes in test-retest reliabilities. In case of non-normal distributions, we calculated both Spearman's rho and the ICC, as Spearman's rho is not affected by non-normality. Spearman's rho did not substantially differ from the ICC. We therefore report only ICCs below.

Results

Structural analyses

The means and standard deviations of volumetric measures in session 1 and session 2 were highly similar. Test-retest reliabilities for both cortical (gray matter and white matter) and subcortical (amygdala and thalamus) volumetric measures were excellent, both across the entire sample (.80 ≤ ICC ≤ .98) and for left-handed (.93 ≤ ICC ≤ .98, with the exception of right amygdala volume ICC = .74) and right-handed (.81 ≤ ICC ≤ .99) participants separately. Differences between left- and right-handed participants in structural test-retest reliabilities were small and not significant, as tested using Fisher's r to *Z* transformation (See Supplementary material for a detailed overview of ICC values).

fMRI test-retest reliability

Table 1 summarizes test-retest reliabilities obtained for maximum cope values. As evident from Table 1, only FFA activity related to face processing (i.e., the contrast face vs. scrambled) showed good reliability, both across the entire sample (ICC = .71), and for left- (ICC = .66) and right-handed (ICC = .75) participants separately. V1 activity showed fair reliability for left-handed participants only (.45 ≤ ICC ≤ .57; except for NeutralUnfamiliar, ICC = .35) and poor reliability for right-handed participants (ICC ≤ .18) as well as across the entire sample (.28 ≤ ICC ≤ .35). Test-retest reliability for the contrast familiar vs. unfamiliar was poor in all tested ROIs (ICC ≤ .28) across the entire sample and for left- and right-handed participants separately, except for FFA activity for left-handed participants, which was fairly reliable (ICC = .46).

As can be seen in Table 1, there is no consistent trend for increasing ICCs with increasing numbers of trials, and reliabilities vary only slightly (.06 ≤ ICC ≤ .60 [78 trials], ICC .07 ≤ ICC ≤ .65 [156 trials], .01 ≤ ICC ≤ .71 [234 trials]). Notably, for V1 we obtained fair reliability for all contrast when analyzing only the first third of the task (78 trials, .40 ≤ ICC ≤ 51), with ICCs seemingly decreasing with increasing numbers of trials, although the differences are small. As shown in Table 1, test-retest reliabilities were somewhat higher for left-handed (.28 ≤ ICC ≤ .66) than for right-handed participants (-.13 ≤ ICC ≤ .75), with the exception of the ICC for FFA activity in the contrast faces vs. scrambled. However, the differences in test-retest-reliability between left- and right-handed participants, as tested using Fisher's r to *Z* transformation, were not statistically significant.

Table 1. Test-retest reliabilities for maximum values of the whole sample for the first third (78 trials N= 42[1]), the first two thirds of the task (156 trials N= 42[1]), and the complete task (234 trials N = 41) , and for left- (N= 20) and right-handed (N= 21) participants separately (234 trials).

ROI	Contrast	Number of trials			Left-handed	Right-handed	Fisher's r to z	
		78	156	234			Z	p
V1	ThreatFamiliar vs. fix	0.41	0.35	0.35	0.57	0.06	1.74	0.08
	ThreatUnfamiliar vs. fix	0.40	0.41	0.35	0.53	0.07	1.53	0.13
	ThreatScrambled vs. fix	0.51	0.40	0.35	0.53	0.13	1.37	0.17
	NeutralFamiliar vs. fix	0.46	0.46	0.29	0.48	0.02	1.51	0.13
	NeutralUnfamiliar vs. fix	0.44	0.41	0.28	0.35	0.18	0.56	0.58
	NeutralScrambled vs. fix	0.43	0.29	0.28	0.45	0.07	1.21	0.23
FFA	Familiar vs. Unfamiliar	0.13	0.25	0.25	0.46	0.04	1.36	0.17
	Face vs. Scrambled	0.60	0.65	0.71	0.66	0.75	-0.52	0.60
IFG	Familiar vs. Unfamiliar	0.06	0.11	0.01	0.28	-0.13	1.22	0.22
STG	Familiar vs. Unfamiliar	0.12	0.07	0.16	0.28	-0.04	0.96	0.34

[1] For one participant, data was only available for the first and the second part of the task, since this participant fell asleep during the third part.

fix = fixation cross

Results obtained for mean and median values are displayed in the Supplementary material (Tables S4.3 and S4.4). Test-retest reliabilities of mean and median values were generally lower than those of maximum values but showed largely the same pattern, with reliable results obtained only for FFA activity related to face processing (face vs. scrambled). We also investigated test-retest reliabilities for maximum values within the anatomically defined ROIs (see Supplementary material, Table S4.2). The ICCs obtained were highly similar to those acquired for the functional masks. In addition, we calculated within-session reliabilities to examine whether low reliability values might be explained by systematically low reliability in one session (see Supplementary material, Table S4.6). For the contrast familiar vs. unfamiliar, we obtained low ICCs for session 1 and session 2 for all tested ROIs. However, V1 activity for the contrasts comparing activity in response to the stimulus conditions to fixation cross, and FFA activity for the contrast face vs. scrambled, tended to be more reliable in session 1 than in session

2, suggesting that some habituation may have occurred between the sessions and/or within session 2. In addition, ICCs for V1 activity were systematically higher for left-handed than for right-handed participants in session 2 (although significant only for the contrast NeutralUnfamiliar vs. fixation cross), but not in session 1. This mainly reflects lower within-session reliability for right-handed participants in session 2 when compared to session 1. In fact, reliability of V1 activity was fair to excellent within session 1 (both across the groups and for left- and right-handed participants separately; ICC ≥ .47) and for left-handed participants within session 2 (ICC ≥ .41) and poor only for right-handed participants within session 2 (.13 ≤ ICC ≤ .31, except NeutralScrambled vs. fixation: ICC = .46). Thus, the habituation effects mentioned above may, in V1, be limited to right-handed participants.

Finally, ICCs for consistency were generally comparable to ICCs obtained for absolute agreement (see Supplementary material, Table S4.5), with the exception of somewhat higher ICCs obtained for V1 activity for left-handed participants for consistency compared to absolute agreement. In addition, ICCs for consistency were systematically higher for left-handed participants than for right-handed participants (although significant only in V1 for the contrasts ThreatFamiliar vs. fixation cross, ThreatScrambled vs. fixation cross, and NeutralScrambled vs. fixation cross after correction for multiple testing).

Discussion

The main purpose of this study was to assess test-retest reliability of significant fMRI activity acquired during a face processing paradigm in a priming context. Reliabilities of structural MRI data were generally excellent, with the exception of good reliability for right amygdala volume measured in left-handed participants. Somewhat lower reliability for amygdala volumes compared to larger subcortical structures were also obtained in earlier research using the same segmentation procedure (Morey et al., 2010). The reliabilities obtained for fMRI data were generally lower than expected. Stable activity was found only for the FFA in response to familiar and unfamiliar faces compared to scrambled faces. In addition, we obtained fairly stable V1 activity in left-handed, but not in right-handed participants. Unexpectedly, adding more trials did not substantially increase test-retest reliability, and in V1, reliability of maximum copes even decreased from fair (ICCs ≥ .40 for 78 trials) to poor (ICCs ≤ .35 for 234 trials) with an increasing number of trials. Regarding handedness, although ICCs for left-handed participants were generally higher than for right-handed participants, the differences were usually not statistically significant (with four exceptions, see Supplementary material). Thus, whether subtle differences in the reliability of fMRI activity exist between left-handed and right-handed individuals remains a topic for investigation. Finally, it is worth noting that we observed larger ICCs for maximum values of activity than for mean and median values. Thus, the results presented in the results section show a relatively optimistic picture.

The good reliability of FFA activity in response to faces compared to scrambled stimuli is in line with earlier research that demonstrated robust changes in FFA activity related to face processing (e.g. Gauthier, Skudlarski, Gore, & Anderson, 2000; Haxby et al., 2001). Our results emphasize that FFA activity reflecting face processing can be measured reliably with fMRI. To the best of our knowledge, only three other studies have investigated the reliability of FFA activity related to face processing. In the first two studies, test-retest reliabilities were fair to good (Sauder et al., 2013) and fair to excellent (Nord, Gray, Charpentier, Robinson, & Roisier, 2017), similar to our reliability estimates. The third study, however, reported low reliability (Lipp et al., 2014). The small sample size in the Lipp et al. (N=14; 2014) study may explain the deviating results, as studies using small sample sizes are at greater risk of drawing incorrect conclusions.

The low reliability of IFG, STG, and FFA activity in response to familiar faces compared to unfamiliar faces may be explained by various factors. As we did not find any significant activity differences between the sessions, the poor ICCs cannot be explained by a significant loss of activity in session 2. However, substantial variation in brain activity over time, even within sessions, within these ROIs may account for low reliability estimates. The low stability of activity differences within each session are in accordance with this interpretation. The 'task' in our study was a free-viewing paradigm. Participants were asked to simply look at the stimuli. As a consequence, we did not control participants' mental processes during the task. Thus, both within- and between-sessions, differences in mental state between participants and within participants over time are possible (e.g. due to variations in attention to and mental operations performed during the task). On the other hand, when participants have to perform a cognitive task during a face processing paradigm, task-specific factors might affect how the brain processes the presented faces which may affect (condition differences) in FFA activity. For instance, in one earlier study, participants were asked to categorize faces for either their gender or their familiarity. In this study, the gender and the familiarity categorization task differentially affected the N170 component of the event-related potential (Goffaux, Jemel, Jacques, Rossion, & Schyns, 2003), a component that has been related to face processing in the fusiform gyrus (Iidaka, Matsumoto, Haneda, Okada, & Sadato, 2006). In future research, including a cognitive task in the face processing paradigm used here may help to focus the attention of the participants on the presented faces. However, unintended effects on the processing of the presented faces by adding such a cognitive task should also be investigated.

Low reliabilities may also be caused by a low signal-to-noise ratio (SNR) of the fMRI data. However, we obtained stable ICCs in the face processing contrast (face vs. scrambled) for FFA activity, and significant effects obtained in session 1 were in accordance with expectations. Therefore, it is unlikely that the low reliabilities were due simply to excessive noise. That ICCs for FFA activity in the contrast faces vs. scrambled stimuli were acceptable whereas ICCs obtained in the FFA, IFG, and the STG for the contrast familiar vs. unfamiliar faces were not may instead be explained by the type of cognitive process reflected in these contrasts. With the contrast faces (unfamiliar and familiar) vs. scrambled stimuli, basic face processing is investigated, and a large number of earlier studies report enhanced FFA activity in response to faces compared to non-facial stimuli (e.g. Gauthier et al., 2000; Haxby et al., 2001). The contrast unfamiliar vs. familiar faces targets the brain processes involved in processing familiarity of the faces presented. Processing face familiarity seems to occur at a later processing stage (Eimer, 2000), and also appears to involve more diverse brain areas, with less consensus across studies concerning the areas involved (Natu & O'Toole, 2011). Nevertheless, changes in FFA, IFG and STG activity in response to familiar faces compared to unfamiliar faces were reported in several earlier studies (see Natu & O'Toole, 2011) in addition to our own (Heckendorf et al., 2016). Thus, although effects may not be as robust as changes in FFA activity related to basic face processing, FFA, IFG, and STG seem to play a role in processing face familiarity. Note also that low ICCs do not necessarily imply that group differences in brain activity in response to different types of stimuli (e.g., differences in response to familiar and unfamiliar faces) cannot be consistently significant. Rather, low ICCs imply that the size of the activity difference for individual participants is not stable over time. In all, more research is needed to further investigate the reliability of significant changes in brain activity related to familiarity processing.

Habituation of brain activity may also decrease ICCs. Because ICCs were consistently low, rather than acceptable when only the first few trials were included in the analyses and not when analyzing the entire task (expected when brain activity habituates within a session) or acceptable for the first but not the second session (expected when habituation occurs between sessions and/or during the second), we did not find strong evidence for habituation effects in IFG, STG and FFA. In contrast, in V1, the lower reliability observed in case of an increasing number of trials as well as the observation of lower within-session reliability for session 2 than session 1 (particularly in right-handers) may reflect habituation effects. Habituation effects in V1 may for instance be caused by repeated exposure to only one unfamiliar-looking and one familiar-looking face per session. Including several familiar-looking and unfamiliar-looking faces may reduce habituation effects. Therefore, future studies may use a face processing paradigm with several different unfamiliar and familiar faces to investigate whether this increases fMRI test-retest reliabilities. Concerning the limited reliability of V1 activity elicited by the six combinations of primes and faces, it should also be noted that it was not possible to model hemodynamic responses to primes and faces separately, because primes and masks were presented very briefly and time-locked to the presentation of the faces. Because, as a consequence, hemodynamic responses to these individual stimuli overlapped substantially, we modelled a single, summed hemodynamic response for the sequence mask-prime-mask-face. In the contrasts with the fixation cross activity in response to faces and primes respectively can therefore not be separated (note that the contrasts faces vs. scrambled and familiar vs. unfamiliar do isolate activity related specifically to processing faces and familiarity as the same primes were presented with each face). In future research trials in which the prime and masks are omitted from the stimulus sequence could be included to enable separate modelling of neural responses to primes and faces, and estimation of the reliability of visual activity in response to primes and faces separately.

In their review, Bennett and Miller (2010) point out that fMRI reliability and the statistical power of fMRI experiments can both be enhanced by adding extra subjects and by increasing the length of the task. However, in our study, adding more trials did not substantially increase reliability. Habituation effects may further compromise the value of added trials. There may thus be a trade-off between an enhanced signal resulting from extra trials, and increased noise with increasing task length due to habituation, subject fatigue or shifts in attention (Bennett & Miller, 2010). Thus, researchers need to carefully weigh these factors when deciding on the number of trials to include in an experiment, and further research investigating what constitutes and which factors affect the optimum number of trials is clearly necessary. Future research may also study habituation itself, by investigating when and how (e.g. at what rate, as a linear or non-linear function of time) habitation takes place, and how to model habituation appropriately in fMRI analysis. This will enable estimation of effects of habituation on the reliability of fMRI data. With regard to differences in fMRI reliability between left-handed and right-handed participants, ICCs tended to be higher for left-handed than right-handed participants, but, apart from four exceptions (see Supplementaries), differences were not statistically significant. As our groups of left-handed and right-handed participants were relatively small, subtle differences in fMRI reliability between left-handed and right-handed individuals warrant attention in future research using larger samples. Note that because we did not find lower reliability for left-handed participants, it seems clear that for tasks like the one used here reliability is no reason to justify the exclusion of left-handed individuals from participation in an fMRI study.

Future research may also address some of the limitations of the current study. First, we examined fMRI reliability for one specific task. Studies examining the reliability of other research paradigms are badly needed. Moreover, future studies could examine fMRI reliability across a range of tasks (e.g. a memory, a motor and a visual task) to increase our understanding of how specific task characteristics may affect fMRI reliability when other relevant parameters, such as the scan procedure, are held constant. Second, participants in our study completed a face processing task embedded in a priming context. Although we did not find significant priming effects, we cannot exclude the possibility that the focus on priming affected test-retest reliability. Reliability studies using a face processing task without priming could confirm that the primes included in our study did not affect the ICCs obtained. Third, with 23 minutes, our task was relatively long. In future studies the task could be split into several runs with short breaks in between, to examine whether this may increase reliabilities.

Finally, we only included female university students in our sample, and the results may therefore not be generalizable to other populations (e.g. men, clinical groups). We chose to focus only on females, because of concerns for sample homogeneity and size. Large-sample studies including, and comparing, both males and females are obviously welcome. Few have investigated fMRI reliability in clinical samples, but the existing studies indicate a lower fMRI reliability in clinical samples compared to healthy controls (see for a review Bennett & Miller, 2010). It would be interesting to use our face processing paradigm in individuals showing aberrant responses to social stimuli such as faces, and examine test-retest reliability of fMRI activity among these individuals. To increase our understanding of the neurological deficits underlying deviant responses to faces, such as those reported in individuals with autism spectrum disorder, it is essential that we can reliably measure face processing in these individuals.

In conclusion, the current study showed relatively low fMRI reliability, with the exception of FFA activity related to face processing. This suggests that the paradigm used in this study, and perhaps fMRI more generally, is not ideally suited to study individual differences in brain activity. Low ICCs for fMRI data seem to be no exception. Although Bennett and Miller (2010) computed an average ICC of .50 across multiple fMRI reliability studies, ICCs varied substantially across ROIs and contrasts examined in individual studies. In addition, poor ICCs were found in several earlier studies focusing on face processing that examined fMRI reliability (e.g. Lipp et al., 2014; Van der Bulk et al., 2013). Also, some of the studies included in the meta-analysis of Bennett & Miller (2010), examined very basic processes, such as motor processes, which has probably led to a higher average ICC. Based on the poor reliability values obtained in our and several other fMRI studies, it is important to look toward factors that may increase the reliability of fMRI measurements. Technical improvements of the MR hardware and software packages used to analyze the MRI data acquired remain desirable to enhance the progress of neuroimaging research. In addition, limited reliability stresses the need for larger samples in fMRI studies, as the associated measurement error in smaller samples elevates the risk of non-reproducible group results. However, larger samples can never compensate for extremely low reliability. Moreover, the fact that the validity of a measurement is limited by its reliability (Shrout, 1998) makes the search for reliable fMRI assessments even more urgent.

References

Bartzokis, G., Mintz, J., Marx, P., Osborn, D., Gutkind, D., Chiang, F., ... & Marder, S. R. (1993). Reliability of in vivo volume measures of hippocampus and other brain structures using MRI. *Magnetic resonance imaging, 11*(7), 993-1006. doi: 10.1016/0730-752X(93)90218-3

Belliveau J.W., Kennedy D.N., Jr, McKinstry R.C., Buchbinder B.R., Weisskoff R.M., Cohen M.S., Vevea J.M., Brady T.J., Rosen B.R (1991). Functional mapping of the human visual cortex by magnetic resonance imaging. *Science, 254*, 716–719.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300.

Bennett, C. M., & Miller, M. B. (2010). How reliable are the results from functional magnetic resonance imaging?. *Annals of the New York Academy of Sciences, 1191*(1), 133-155. doi: 10.1111/j.1749-6632.2010.05446.x

Beyers, W., and Goossens, L. (2003). Psychological separation and adjustment to university: moderating effects of gender, age and perceived parenting style. *J. Adolesc. Res.* 18, 363–382. doi: 10.1177/0743558403018004003

Birn, R. M., Molloy, E. K., Patriat, R., Parker, T., Meier, T. B., Kirk, G. R., ... & Prabhakaran, V. (2013). The effect of scan length on the reliability of resting-state fMRI connectivity estimates. *Neuroimage, 83*, 550-558. doi: 10.1016/j.neuroimage.2013.05.099

Brandt, D. J., Sommer, J., Krach, S., Bedenbender, J., Kircher, T., Paulus, F. M., & Jansen, A. (2013). Test-retest reliability of fMRI brain activity during memory encoding. *Frontiers in psychiatry, 4*, 163. doi: 10.3389/fpsyt.2013.00163

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365-376. doi:10.1038/nrn3475

Cao, H., Plichta, M. M., Schäfer, A., Haddad, L., Grimm, O., Schneider, M., ... & Tost, H. (2014). Test–retest reliability of fMRI-based graph theoretical properties during working memory, emotion processing, and resting state. *Neuroimage, 84*, 888-900. doi: 10.1016/j.neuroimage.2013.09.013

Chen, E. E., & Small, S. L. (2007). Test–retest reliability in fMRI of language: group and task effects. *Brain and language, 102*(2), 176-185. doi: 10.1016/j.bandl.2006.04.015

Convit, A., McHugh, P., Wolf, O. T., De Leon, M. J., Bobinski, M., De Santi, S., ... & Tsui, W. (1999). MRI volume of the amygdala: a reliable method allowing separation from the hippocampal formation. *Psychiatry Research: Neuroimaging, 90*(2), 113-123.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology, 10*, 85.

De Corte, K., Buysse, A., Verhofstadt, L. L., Roeyers, H., Ponnet, K., and Davis, M. H. (2007). Measuring empathic tendencies: reliability and validity of the Dutch version of the interpersonal reactivity index. *Psychol. Belg.* 47, 235–260.

Eimer, M. (2000). Attentional modulations of event-related brain potentials sensitive to faces. *Cognitive Neuropsychology, 17*(1-3), 103-116. doi: 10.1080/026432900380517

Engell A.D. and McCarthy G. (2013). fMRI activation by face and biological motion perception: Comparison of response maps and creation of probabilistic atlases. *NeuroImage, 74,* 140-151. doi: 10.1016/j.neuroimage.2013.02.025

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.). Washington, DC: The American Council on Education and the National Council on Measurement in Education.

Fleiss, J. L. (1986). Reliability of measurement. In *The Design and Analysis of Clinical Experiments*. New York : Wiley. doi: 10.1002/9781118032923.ch1

Fliessbach, K., Rohe, T., Linder, N. S., Trautner, P., Elger, C. E., & Weber, B. (2010). Retest reliability of reward-related BOLD signals. *Neuroimage, 50*(3), 1168-1176. doi: 10.1016/j.neuroimage.2010.01.036

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature neuroscience, 3*(2), 191-197. doi: 10.1038/72140

Gay, L.R. (1987). "Selection of measurement instruments". In *Educational Research: Competencies for Analysis and Application* (3rd ed.). New York: Macmillan.

Goffaux, V., Jemel, B., Jacques, C., Rossion, B., & Schyns, P. G. (2003). ERP evidence for task modulations on face perceptual processing at different spatial scales. *Cognitive Science, 27*(2), 313-325. doi: 10.1207/s15516709cog2702_8

Gorgolewski, K. J., Storkey, A. J., Bastin, M. E., Whittle, I., & Pernet, C. (2013). Single subject fMRI test–retest reliability metrics and confounding factors. *Neuroimage, 69*, 231-243. doi: 10.1016/j.neuroimage.2012.10.085

Harrington, G. S., Farias, S. T., Buonocore, M. H., & Yonelinas, A. P. (2006). The intersubject and intrasubject reproducibility of FMRI activation during three encoding tasks: implications for clinical applications. *Neuroradiology, 48*(7), 495-505. doi: 10.1007/s00234-006-0083-2

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425-2430. doi: 10.1126/science.1063736

Heckendorf, E., Huffmeijer, R., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2016). Neural Processing of Familiar and Unfamiliar Children's Faces: Effects of Experienced Love Withdrawal, but No Effects of Neutral and Threatening Priming. *Frontiers in Human Neuroscience*. doi: 10.3389/fnhum.2016.00231

Huffmeijer, R., Bakermans-Kranenburg, M. J., Alink, L. R., & van IJzendoorn, M. H. (2014). Reliability of event-related potentials: the influence of number of trials and electrodes. *Physiology & behavior, 130,* 13-22. doi:10.1016/j.physbeh.2014.03.008

Iidaka, T., Matsumoto, A., Haneda, K., Okada, T., & Sadato, N. (2006). Hemodynamic and electrophysiological relationship involved in human face processing: evidence from a combined fMRI-ERP study. *Brain and Cognition, 60,* 176-186. doi: 10.1016/j.bandc.2005.11.004

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage, 17*(2), 825-841. doi: 10.1006/nimg.2002.1132

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., and Smith, S. M. (2012). FSL. *Neuroimage 62,* 782–790. doi: 10.1016/j.neuroimage.2011.09.015

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience, 17*(11), 4302-4311.

Lipp, I., Murphy, K., Wise, R. G., & Caseras, X. (2014). Understanding the contribution of neural and physiological signal variation to the low repeatability of emotion-induced BOLD responses. *Neuroimage, 86,* 335-342. doi: 10.1016/j.neuroimage.2013.10.015

McManus, I. C. (2009). The history and geography of human handedness.*Language lateralization and psychosis,* 37-57.

Morey, R. A., Selgrade, E. S., Wagner, H. R., Huettel, S. A., Wang, L., & McCarthy, G. (2010). Scan-rescan reliability of subcortical brain volumes derived from automated segmentation. *Human brain mapping, 31*(11), 1751-1762.

Natu, V., & O'Toole, A. J. (2011). The neural processing of familiar and unfamiliar faces: A review and synopsis. *British Journal of Psychology, 102*(4), 726-747. doi: 10.1111/j.2044-8295.2011.02053.x

Nord, C. L., Gray, A., Charpentier, C. J., Robinson, O. J., & Roiser, J. P. (2017). Unreliability of putative fMRI biomarkers during emotional face processing. *NeuroImage, 156,* 119-127.

Nummenmaa, L., Hirvonen, J., Parkkola, R., & Hietanen, J. K. (2008). Is emotional contagion special? An fMRI study on neural systems for affective and cognitive empathy. *Neuroimage, 43*(3), 571-580. doi: 10.1016/j.neuroimage.2008.08.014

Plichta, M. M., Schwarz, A. J., Grimm, O., Morgen, K., Mier, D., Haddad, L., ... & Colman, P. (2012). Test–retest reliability of evoked BOLD signals from a cognitive–emotive fMRI test battery. *Neuroimage, 60*(3), 1746-1758. doi: 0.1016/j.neuroimage.2012.01.129

Puce, A., Allison, T., Gore, J. C., & McCarthy, G. (1995). Face-sensitive regions in human extrastriate cortex studied by functional MRI. *Journal of neurophysiology, 74*(3), 1192-1199.

Sauder, C. L., Hajcak, G., Angstadt, M., & Phan, K. L. (2013). Test-retest reliability of amygdala response to emotional faces. *Psychophysiology, 50*(11), 1147-1156. doi: 10.1111/psyp.12129

Schacher, M., Haemmerle, B., Woermann, F. G., Okujava, M., Huber, D., Grunwald, T., ... & Jokeit, H. (2006). Amygdala fMRI lateralizes temporal lobe epilepsy. *Neurology, 66*(1), 81-87. doi: 10.1212/01.wnl.0000191303.91188.00

Schludermann, S., and Schludermann, E. (1983). Sociocultural change and adolescents perceptions of parent behavior. *Dev. Psychol.* 19, 674–685. doi: 10.1037/0012-1649.19.5.674

Sergent J, Ohta S, MacDonald B. (1992). Functional neuroanatomy of face and object processing. A positron emission tomography study. *Brain, 115*(1), 15-36. doi: 10.1093/brain/115.1.15

Shrout, P. E. (1998). Measurement reliability and agreement in psychiatry. *Statistical methods in medical research*, 7(3), 301-317.

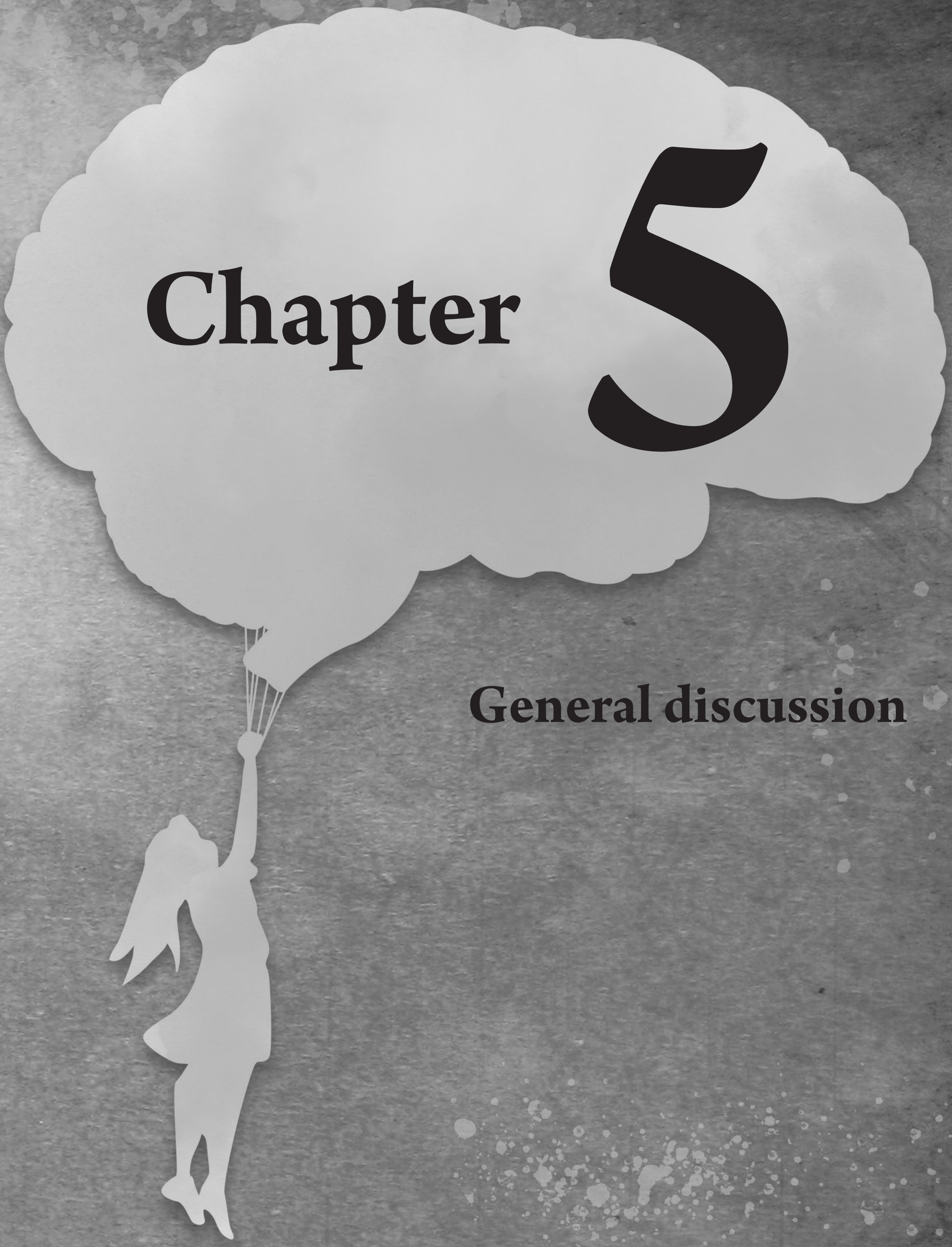
Smith, S. M. (2002). Fast robust automated brain extraction. *Human brain mapping, 17*(3), 143-155. doi: 10.1002/hbm.10062

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., ... & Niazy, R. K. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage, 23,* S208-S219. doi: 10.1016/j.neuroimage.2004.07.051

van den Bulk, B. G., Koolschijn, P. C. M., Meens, P. H., van Lang, N. D., van der Wee, N. J., Rombouts, S. A., ... & Crone, E. A. (2013). How stable is activation in the amygdala and prefrontal cortex in adolescence? A study of emotional face processing across three measurements. *Developmental cognitive neuroscience, 4,* 65-76. doi: 10.1016/j.dcn.2012.09.005

Van Strien, J. W. (1992). Classificatie van links-en rechtshandige proefpersonen. *Nederlands tijdschrift voor de psychologie, 47,* 88-92.

Willems, R. M., Van der Haegen, L., Fisher, S. E., & Francks, C. (2014). On the other hand: including left-handers in cognitive neuroscience and neurogenetics. *Nature Reviews Neuroscience, 15*(3), 193-201. doi: 10.1038/nrn3679



Chapter 5

General discussion

The main aim of the research presented in this thesis was to increase our insight into individual differences in the neural processing and appraisals of child faces with varying degrees of facial resemblance with the participant. The central question we tried to answer in this dissertation was whether individual differences in experiences of love withdrawal, empathy and FFA activity, and the context in which child faces are shown, may influence reactions to and appraisal of self-resembling child faces. The results obtained in Chapter 2 and 3 of this thesis show differences in the (neural) processing of self-resembling and non-resembling child faces, and effects of facial resemblance on participants' negative appraisals of children's faces (i.e. less negative appraisals with increasing facial resemblance). Love withdrawal moderated the effects of facial resemblance on neural processing and negative appraisals of children's faces. We did not obtain any effects involving threat priming or empathy. In Chapter 4, we analyzed the test-retest reliability of the fMRI data acquired with our face processing task. Generally, we acquired lower reliabilities than might have been expected, with the exception of good test-retest reliability of FFA activity in response to familiar and unfamiliar faces compared to scrambled faces. Below, we will elaborate in more detail on the results. Limitations, implications and directions for future research will be discussed.

Facial resemblance: Effects on neural processing and evaluations of child faces

We found differences in the neural processing of child faces that resembled the participant's face compared to child faces that did not. Self-resembling child faces evoked enhanced activity in brain areas related to self-referential processing and emotion-regulation (e.g. IFG, MFG; insular; see for a review Devue & Brédart, 2011). These processes have also been shown to play a role in parental behavior (Swain et al., 2014). In addition, child faces that resembled the participant's face also evoked greater activity in some core face processing areas, such as the lateral fusiform gyrus, compared to child faces that did not resemble the participant's face, suggesting preferential processing of self-resembling child faces (Natu & O'Toole, 2011).

Child faces that resembled the participants more were also evaluated less negatively. Decreasing negative evaluations with increasing self-resemblance of child faces may promote caregiving behaviors directed at close genetic relatives, since, on average, closely related individuals resemble each other more than non-related individuals (Alvergne, Faurie, & Raymond, 2007; Bressan & Grassi, 2004; DeBruine et al., 2009; Maloney & Dal Martello, 2006). In particular, parental caregiving behaviors directed at biological children may be enhanced due to reduced negative appraisals experienced by mothers and fathers when viewing their biological children (that facially resemble themselves) compared to negative appraisals experienced when viewing other, non-related children (that do not facially resemble themselves). Moreover, neural activity was also associated with participants' negative appraisals: Higher activity in the fusiform face area (FFA), in combination with higher levels of love-withdrawal, was related to a stronger decrease in negative appraisals of child faces that resembled the participants more. Thus, more extensive face processing (reflected by enhanced FFA activity) was associated with larger effects of facial resemblance on the negative appraisals of children's faces. The FFA is involved in the processing of the identity cues of faces, such as facial resemblance, in order to discriminate between individual faces (see for a review Kanwisher & Yovel, 2006). Higher FFA activity in the current study may therefore reflect an enhanced sensitivity to facial resemblance, which may explain the larger effects of facial resemblance on (negative) appraisals of children's faces observed in participants showing higher FFA activity.

In contrast, brain areas associated with theory of mind, such as bilateral superior temporal gyrus, showed enhanced activity in response to non-resembling child faces. Theory of mind refers to the ability to assign mental states, such as beliefs, desires and intentions, to others (Frith & Frith, 1999; Schurz, Radau, Aichhorn, Richlan, & Perner, 2014). In a previous study, in addition to deactivations in some other brain regions, decreased activity in the superior temporal sulcus was obtained when mothers viewed pictures of their own children compared to pictures of familiar, but non-related children (Bartels & Zeki, 2004). It was suggested that maternal attachment may deactivate brain regions involved in social validation (e.g. estimating the mental states of others), since, compared to less or non-familiar people, individuals may feel less need to determine the mental states of people they are highly familiar with. In the studies presented in this thesis, we used pictures of child faces that resembled the participant's face. The association of facial resemblance with close genetic relatedness may explain why we obtained deactivations in response to self-resembling child faces in similar brain regions as Bartels & Zeki (2004) found when mothers viewed pictures of their own children.

In addition, people may assume that individuals that facially resemble themselves (external resemblance) also have similar desires, intentions, and beliefs (internal resemblance) as themselves. As a consequence, they may feel less need to thoroughly investigate the mental states of self-resembling individuals, compared to non-resembling persons, which may also explain the deactivation of bilateral superior temporal gyrus in response to self-resembling child faces obtained in our study. In conclusion, the results indicate that the degree of resemblance of child faces with the participant's face affects both participants' (neural) processing and their (negative) appraisals of child faces.

Individual differences and context

Parental love withdrawal moderated brain activity in face processing areas (in particular occipital fusiform gyrus). We obtained greater differences in neural responses to self-resembling and non-resembling child faces in participants who reported more love withdrawal. In addition, experiences of love withdrawal also influenced participants' negative evaluations of child faces with differing degrees of self-resemblance: Stronger effects of facial resemblance (i.e. stronger decreases in negative evaluations) were visible in participants showing high FFA activity and reporting high love withdrawal. The results indicate that effects of childhood experiences of love withdrawal may still be present in early adulthood, as evidenced by the effects of love withdrawal on participants' neural responses and evaluations of the child faces.

Our results are in line with two earlier EEG-studies, in which effects of love withdrawal on the neural processing of emotional faces were observed in young female adults. In these studies, experiences of maternal love withdrawal were related to enhanced amplitudes of the Vertex Positive Potential in response to emotional faces (VPP; Huffmeijer, Tops, Alink, Bakermans-Kranenburg, & Van IJzendoorn, 2011; Huffmeijer et al., 2013). The VPP is an event-related potential (ERP) component associated with processing the configuration of faces, with larger amplitudes indicating more extensive face processing (Luo et al., 2010; Rossion et al., 1999). The VPP is thought to reflect the same neural activity as the N170 (Jemel et al., 2003; Itier & Taylor, 2002; Joyce & Rossion, 2005), which has been directly related to activity in the fusiform gyrus (Iidaka, Matsumoto, Haneda, Okdada, & Sadato, 2006). Thus, both the results presented in this thesis and the results of previous EEG studies indicate that childhood experiences of love withdrawal may affect how individuals process faces. Importantly, early experiences of love withdrawal seem to affect the identification of configural cues, both those that communicate identity and those that communicate emotional state.

In Chapter 2 of this thesis, by assessing participants' reactions to self-resembling child faces that were preceded by primes depicting threatening scenes, we aimed to study the neural processes associated with protective caregiving reactions. However, we did not obtain significant priming effects. We discussed various explanations for the absence of priming effects: Preconscious processing may not have occurred because of the complexity of the primes, which may have impeded rapid extraction of threatening and neutral context. Or, alternatively, the threatening scenes included as primes may not have evoked protective reactions in the participants. However, the non-significant priming effects were also in line with the results of several previous studies that questioned the existence of a threat processing system in the human brain that processes threat-related stimuli without conscious awareness (Pessoa & Adolphs, 2010; Hoffmann, Lipka, Mothes-Lasch, Miltner, & Straube, 2012). The results of earlier research also suggest that primes included in some previous studies were not really subliminal, which may explain why some earlier studies obtained significant priming effects (Pessoa, Japee & Ungerleider, 2005). Nonetheless, before we can draw definite conclusions concerning the (non-)existence of preconscious threat processing, more research should be conducted on the effects of threat priming.

Test-retest reliability

It is crucial to examine the test-retest reliability of fMRI experiments, since the validity of a finding is limited by its reliability (Feldt & Brennan, 1989; Gay, 1987). Unfortunately, fMRI test-retest reliability studies are relatively scarce. In Chapter 4, we presented our results with regard to the test-retest reliability of the face processing task included in this thesis to increase our understanding of fMRI reliability. We obtained good test-retest reliability for the face processing contrast (FFA: faces vs. scrambled stimuli), indicating that changes in FFA activity related to face processing can be measured reliably with fMRI. However, test-retest reliabilities for processing familiarity (contrast: familiar vs. unfamiliar faces) were low in all tested regions of interest, and basic visual processing of the stimuli in primary visual cortex (V1) was only modestly reliable.

The stable activity obtained in the FFA suggests that changes in FFA activity in response to faces compared to non-facial stimuli can be used to investigate individual differences in face processing, as we did in Chapter 3 of this thesis. Changes in brain activity observed in response to faces vs. scrambled stimuli reflect an early processing stage (Eimer, 2000) that has robustly been located to a specific area within the fusiform gyrus, accordingly named the fusiform face area ([FFA] Gauthier, Skudlarski, Gore, & Anderson, 2000; Haxby et al., 2001; Gobbini & Haxby, 2007). Processing familiarity (contrast: familiar vs. unfamiliar faces) on the other hand seems to occur at a later processing stage, involving a more distributed network of brain areas associated with several cognitive and affective processes (Natu & O'Toole, 2011). If replicated, our findings would suggest that activity changes in highly localized brain areas during early processing stages (i.e. identifying a facial stimulus) can be measured more reliably than changes in brain activity occurring at later processing stages which are related to more subtle differences between stimuli (i.e. identifying differences in the degree of familiarity of two faces).

Low intra-class-correlations (ICCs) imply that, over time, the size of activity measured in individual participants is not stable. Low ICCs are particularly problematic for studies investigating individual differences. However, low stability of individual differences not necessarily implies that group differences cannot be investigated. In view of the limited reliability, our results presented in Chapter 2 of this thesis require replication. Future studies may prove that familiarity processing cannot be reliably measured with fMRI. However, future research could also indicate that specific features of our face processing task (e.g. combining primes depicting threatening scenes with pictures of self-resembling faces) have negatively affected the ICCs for familiarity processing in the current study.

We also examined possible differences in test-retest reliability between left- and right-handed participants. We generally did not obtain statistically significant reliability differences between left- and right-handed participants. Thus, in fMRI research that uses tasks similar to the task that we used, reliability concerns are no reason to exclude left-handed participants. Finally, we investigated whether increasing the number of trials of our research paradigm would lead to enhanced ICCs. Surprisingly, ICCs did not consistently increase with an increasing number of trials, suggesting that adding more trials in fMRI research may not always increase test-retest reliabilities. Our results indicate that researchers should carefully balance the advantages and disadvantages of adding more trials to a research paradigm. Longer tasks provide researchers with more data points, resulting in an enhanced signal. However, increasing the length of a task may also be associated with lower attention and arousal of participants, increased fatigue, and habituation effects, thereby increasing noise with increasing task-length (Bennett & Miller, 2010).

Low test-retest reliabilities are not uncommon in fMRI research (Bennett & Miller, 2010). Researchers should therefore always aim to optimize the reliability of their fMRI data. However, since fMRI research is expensive, including two sessions to examine the test-retest reliability of a research paradigm as a standard practice may not be feasible for most fMRI studies. Alternatively, researchers could investigate the split-half reliability of their fMRI data, as within-session reliability also gives an indication of the reliability of the acquired data.

Limitations and future directions

Future research could also address some of the limitations of the studies described in the current thesis. First of all, our results are based on a sample of young female adults. We aimed to provide a ‘proof of concept’ in a homogenous sample to gain first insights in the neural correlates of (maternal) protective behavior. However, it is possible that the inclusion of only young adult participants without children of their own has influenced the findings. Therefore, the research presented in this thesis should be extended to samples of mothers who will be viewing pictures of their children in the MRI scanner. In addition, we used self-report questionnaires to examine experiences of love withdrawal and participants’ levels of empathy, although the accuracy and reliability of participants’ (retrospective) self-reports is generally limited. In future studies, other (non-retrospective) measures of experiences of love withdrawal could be acquired in longitudinal research. In addition, instead of relying on participants’ self-reports, observations could be used to examine participants’ levels of empathy. In future research, other neuroimaging techniques, such as EEG, could be used to examine individual differences between neural responses to self-resembling and non-resembling child faces. Two previous studies that examined test-retest reliability of ERP components evoked by faces acquired excellent reliability for ERP amplitudes evoked by emotional faces (Cassidy, Robertson, & O’Connell, 2012; Huffmeijer, Bakermans-Kranenburg, Alink, & Van IJzendoorn, 2014) Future studies could for instance use ERPs to examine individual differences in reactions to child faces with differing degrees of resemblance with the participant's face.

Finally, differences in test-retest reliability between left- and right-handed participants were examined in relatively small subgroups (left-handed N=20, right-handed N=21) and future research should include larger samples for this comparison. In general, it seems important to invest in fMRI test-retest reliability studies with larger samples. In a previous meta-analysis, the average sample size of the included fMRI reliability studies was only 11 participants (Bennett & Miller, 2010), indicating that the power of most fMRI reliability studies was very low. More fMRI test-retest reliability studies with larger samples and various research paradigms are therefore badly needed before we can draw conclusions concerning the reliability of fMRI data. Low sample sizes are not only an issue in fMRI reliability studies, but in fMRI research in general (Poldrack et al., 2017), which may lead to low reproducibility of the findings. In low-powered studies, the chance to obtain a significant result is greatly reduced. At the same time, the chance that significant findings reflect false positives is substantially increased (Button et al., 2013). In an earlier attempt to replicate findings obtained for brain structure and behavior associations (i.e. relating brain structure to participants' scores on a psychological or behavioral trait), the replication rate was very low (Boekel, 2015). Low reproducibility of findings is evident throughout biomedical and psychological research, and concerns were raised with regard to the reliability and validity of both behavioral and biological measures (Begley & Ellis, 2012; Klein et al., 2014; Open Science Collaboration, 2015; Prinz, Schlange & Asadullah, 2011). Thus, concerns with regard to the reliability and reproducibility of findings seem to be a more general issue of the current state of science. To increase the credibility of current science, and of fMRI research in particular, large-sampled, high quality research should therefore be strongly encouraged.

Conclusion

In this thesis, we investigated the relation between participants' brain activity, their evaluations of child faces with differing degrees of resemblance with the participant's face, and participants' experiences of love withdrawal, with the broader aim of increasing our understanding of the interplay between brain and behavior. Examining the link between brain activity and behavior is essential to expand our knowledge about how changes in brain activity are expressed in individuals' observable behavior, thoughts and feelings. By investigating the effects of individuals' early experiences on changes in the neural processing of visual and auditory stimuli, we can increase our understanding of long-lasting effects early (negative and positive) experiences may have on later brain functioning. Ideally, longitudinal studies should address the long-term effects of experiences of love withdrawal on the neural processing of external stimuli, and on individuals' observable behavior, thoughts and feelings. The current thesis also emphasizes the importance of test-retest reliability studies in fMRI research. To increase our confidence in the findings of fMRI research, investments in reliability and replication studies should be encouraged. In addition, investments in high-quality research with sufficient sample sizes are crucial for further progress in understanding human brain functioning.

References

Alvergne, A., Faurie, C., & Raymond, M. (2007). Differential facial resemblance of young children to their parents: who do children look like more?. *Evolution and Human behavior, 28*(2), 135-144. doi: 10.1016/j.evolhumbehav.2006.08.008

Bartels, A., & Zeki, S. (2004). The neural correlates of maternal and romantic love. *Neuroimage, 21*(3), 1155-1166. doi: 0.1016/j.neuroimage.2003.11.003

Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature, 483*(7391), 531-533. doi: 10.1038/483531a

Boekel, W., Wagenmakers, E. J., Belay, L., Verhagen, J., Brown, S., & Forstmann, B. U. (2015). A purely confirmatory replication study of structural brain-behavior correlations. *Cortex, 66*, 115-133. doi: 10.1016/j.cortex.2014.11.019

Bressan, P., & Grassi, M. (2004). Parental resemblance in 1-year-olds and the Gaussiancurve. *Evolution and Human Behavior, 25*(3), 133-141. doi: 10.1016/j.evolhumbehav.2004.03.001

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365-376. doi:10.1038/nrn3475

Cassidy, S. M., Robertson, I. H., & O'Connell, R. G. (2012). Retest reliability of event-related potentials: Evidence from a variety of paradigms. *Psychophysiology, 49*(5), 659-664. doi: 10.1111/j.1469-8986.2011.01349.x

DeBruine, L. M., Smith, F. G., Jones, B. C., Roberts, S. C., Petrie, M., & Spector, T. D. (2009). Kin recognition signals in adult faces. *Vision research, 49*(1), 38-43. doi: 10.1016/j.visres.2008.09.025

Devue, C., and Brédart, S. (2011). The neural correlates of visual self-recognition. *Consciousness And Cognition: An International Journal, 20*(1), 40-51. doi:10.1016/j.concog.2010.09.007

Eimer, M. (2000). Attentional modulations of event-related brain potentials sensitive to faces. *Cognitive Neuropsychology, 17*(1-3), 103-116. doi: 10.1080/026432900380517

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.). Washington, DC: The American Council on Education and the National Council on Measurement in Education.

Frith, C.D., and Frith, U. (1999). Interacting minds: A biological basis. *Science, 286,* 1692 – 1695. doi:10.1126/science.286.5445.1692

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature neuroscience, 3*(2), 191-197. doi: 10.1038/72140

Gay, L.R. (1987). "Selection of measurement instruments". In *Educational Research: Competencies for Analysis and Application* (3rd ed.). New York: Macmillan.

Gobbini, M. I., & Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia, 45*(1), 32-41. doi: 10.1016/j.neuropsychologia.2006.04.015

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425-2430. doi: 10.1126/science.1063736

Hoffmann, M., Lipka, J., Mothes-Lasch, M., Miltner, W. H., & Straube, T. (2012). Awareness modulates responses of the amygdala and the visual cortex to highly arousing visual threat. *NeuroImage, 62*(3), 1439-1444. doi: 10.1016/j.neuroimage.2012.05.064

Huffmeijer, R., Alink, L. R., Tops, M., Grewen, K. M., Light, K. C., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2013). The impact of oxytocin administration and maternal love withdrawal on event-related potential (ERP) responses to emotional faces with performance feedback. *Hormones and behavior, 63*(3), 399-410. doi: 10.1016/j.yhbeh.2012.11.008

Huffmeijer, R., Bakermans-Kranenburg, M. J., Alink, L. R., & van IJzendoorn, M. H. (2014). Reliability of event-related potentials: the influence of number of trials and electrodes. *Physiology & behavior, 130*, 13-22. doi: 10.1016/j.physbeh.2014.03.008

Huffmeijer, R., Tops, M., Alink, L. R., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2011). Love withdrawal is related to heightened processing of faces with emotional expressions and incongruent emotional feedback: evidence from ERPs. *Biological Psychology, 86*(3), 307-313. doi: 10.1016/j.biopsycho.2011.01.003

Iidaka, T., Matsumoto, A., Haneda, K., Okada, T., & Sadato, N. (2006). Hemodynamic and electrophysiological relationship involved in human face processing: Evidence from a combined fMRI–ERP study. *Brain and cognition, 60*(2), 176-186. doi: 10.1016/j.bandc.2005.11.004

Itier, R. J., & Taylor, M. J. (2002). Inversion and contrast polarity reversal affect both encoding and recognition processes of unfamiliar faces: a repetition study using ERPs. *Neuroimage, 15*(2), 353-372. doi: 10.1006/nimg.2001.0982

Jemel, B., Schuller, A. M., Cheref-Khan, Y., Goffaux, V., Crommelinck, M., & Bruyer, R. (2003). Stepwise emergence of the face-sensitive N170 event-related potential component. *Neuroreport, 14*(16), 2035-2039. doi: 10.1097/00001756-200311140-00006

Joyce, C., & Rossion, B. (2005). The face-sensitive N170 and VPP components manifest the same brain processes: the effect of reference electrode site. *Clinical Neurophysiology, 116*(11), 2613-2631. doi: 10.1097/00001756-200311140-00006

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 361*(1476), 2109-2128.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams Jr, R. B., Bahník, Š., Bernstein, M. J., ... & Cemalcilar, Z. (2014). Investigating variation in replicability. *Social psychology*. doi: 10.1027/1864-9335/a000178

Luo, W., Feng, W., He, W., Wang, N. Y., & Luo, Y. J. (2010). Three stages of facial expression processing: ERP study with rapid serial visual presentation. *Neuroimage, 49*(2), 1857-1867. doi: 10.1016/j.neuroimage.2009.09.018

Maloney, L. T., & Dal Martello, M. F. (2006). Kin recognition and the perceived facial similarity of children. *Journal of Vision, 6*(10), 4-4. doi:10.1167/6.10.4

Natu, V., & O'Toole, A. J. (2011). The neural processing of familiar and unfamiliar faces: A review and synopsis. *British Journal of Psychology, 102*(4), 726-747. doi: 10.1111/j.2044-8295.2011.02053.x

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), 1-8.

Pessoa, L., and Adolphs, R. (2010). Emotion processing and the amygdala: From a 'low roadto 'many roads' of evaluation biological significance. *Nature Reviews Neuroscience, 11*(11), 7730782. doi:10.1038/nrn2920

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., ... & Yarkoni, T. (2017). Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience, 18*(2), 115-126. doi:10.1038/nrn.2016.167

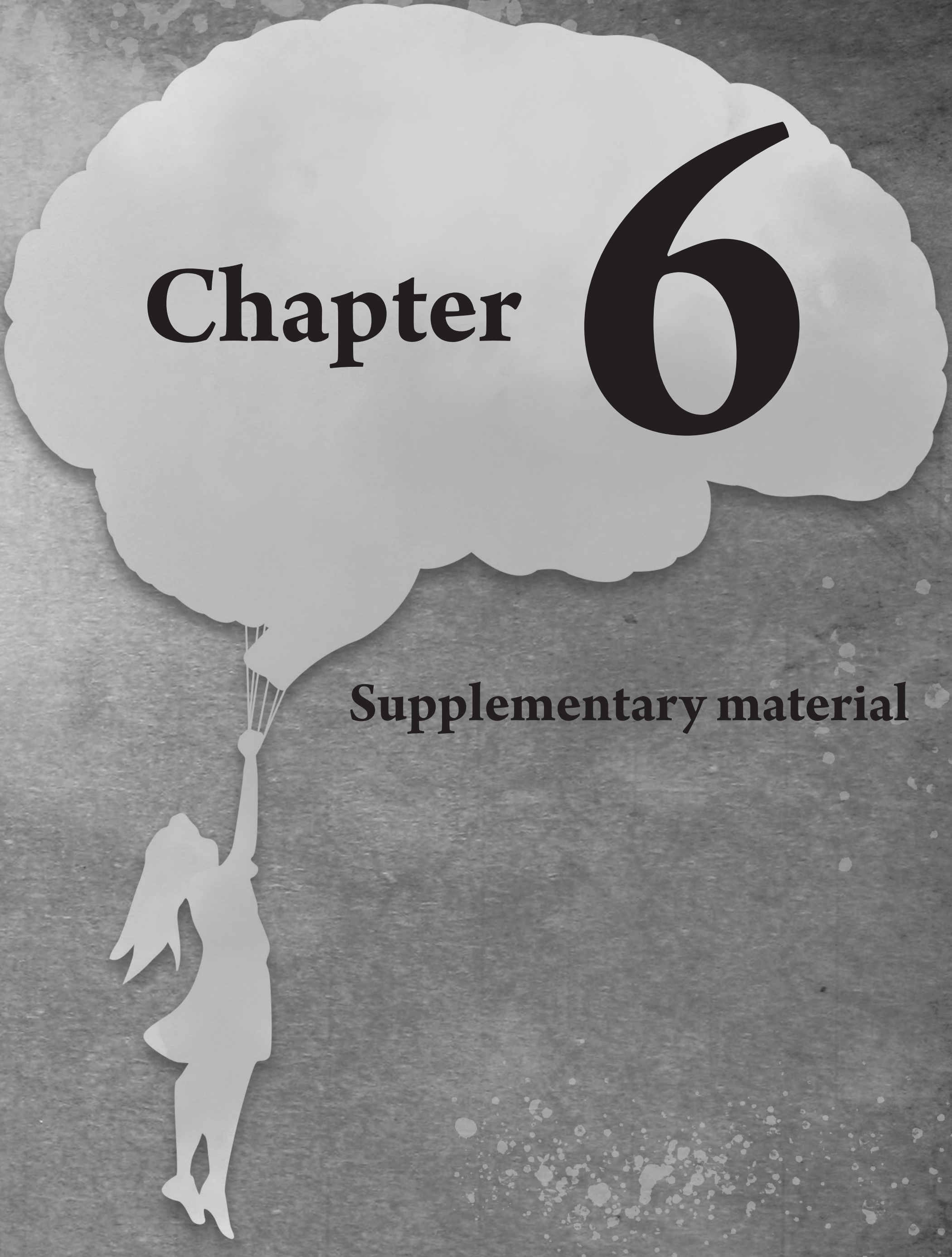
Pessoa, L., Japee, S., & Ungerleider, L. G. (2005). Visual awareness and the detection of fearful faces. *Emotion, 5*(2), 243. doi: 10.1037/1528-3542.5.2.243

Prinz, F., Schlange, T., & Asadullah, K. (2011). Believe it or not: how much can we rely on published data on potential drug targets?. *Nature reviews Drug discovery, 10*(9), 712-712. doi:10.1038/nrd3439-c1

Rossion, B., Delvenne, J. F., Debatisse, D., Goffaux, V., Bruyer, R., Crommelinck, M., & Guérit, J. M. (1999). Spatio-temporal localization of the face inversion effect: an event-related potentials study. *Biological psychology, 50*(3), 173-189. doi: 10.1016/S0301-0511(99)00013-7

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., and Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience And Biobehavioral Review,* 429-34, doi:10.1016/j.neubiorev.2014.01.009

Swain, J. E., Kim, P., Spicer, J., Ho, S. S., Dayton, C. J., Elmadih, A., & Abel, K. M. (2014). Approaching the biology of human parental attachment: Brain imaging, oxytocin and coordinated assessments of mothers and fathers. *Brain Research, 158078*-101. doi:10.1016/j.brainres.2014.03.007



Chapter 6

Supplementary material

Supplementary material Chapter 2

Preliminary analyses

To confirm that our facial stimuli reliably activated face processing areas, we contrasted brain activity in response to familiar-looking and unfamiliar-looking faces with brain activity in response to scrambled faces. Group means were tested using one-sample t-tests. All statistical images were thresholded using clusters determined by $Z > 2.3$ (t-values are automatically converted to z-statistics) and a cluster-corrected significance threshold of $p < 0.05$ (Worsley, 2001). As shown in Table S2.1, faces elicited heightened brain activity in bilateral occipital and temporal areas known to be particularly involved in face processing, including the infero-lateral occipital cortex and occipito-temporal fusiform gyrus (clusters 3 and 4). In addition, faces elicited heightened activity in bilateral intracalcarine and supracalcarine cortex (cluster 2), and in several right hemisphere areas that are part of the brain's socio-emotional networks, including a cluster encompassing parts of the MFG, IFG, insular cortex, and precentral gyrus (cluster 6), as well as in a right occipito-parietal cluster (cluster 5, including the superior lateral occipital cortex, and angular gyrus, extending into superior parietal areas), and in a cluster including parts of the right orbitofrontal cortex, amygdala, putamen, and brain stem (cluster 1).

Table S2.1. MNI coordinates and Z-max values for clusters significantly activated in response to unfamiliar- and familiar-looking faces compared to scrambled stimuli (contrast: face > scrambled).

Cluster	Size	Region	*Z-max*	MNI coordinates for Z-max		
				x	y	z
6	4369	Right MFG and IFG	5.12	48	30	20
5	2746	Right superior parietal lobe	5.25	36	-52	46
4	2576	Right temporal occipital fusiform	5.64	44	-52	-22
3	1381	Left infero-lateral occipital cortex	6.29	36	-82	-2
2	1188	Bilateral intracalcarine cortex	5.06	-14	-64	6
1	1073	Right amygdala	4.18	14	-6	-16

Supplementary material Chapter 4

Test-retest reliability structural MRI

Table S4.1. MRI volumetric measures of the structural analyses for the whole sample (N = 41) and test-retest reliability for the whole sample and for left-handed (N = 20) and right-handed (N = 21) participants separately.

Area	Volume s1 in ml		Volume s2 in ml		ICCs			Fisher's r to z	
	Mean	SD	Mean	SD	Total	Left-handed	Right-handed	Z	p
Gray matter	585.06	41.88	587.30	42.14	0.98	0.96	0.99	-1.68	0.09
White matter	617.67	49.57	614.40	48.08	0.97	0.98	0.97	0.87	0.38
Left amygdala	1.40	0.15	1.41	0.16	0.88	0.94	0.81	1.84	0.07
Right amygdala	1.29	0.20	1.28	0.24	0.80	0.74	0.85	-0.85	0.40
Left thalamus	8.06	0.58	8.07	0.54	0.96	0.97	0.95	0.73	0.47
Right thalamus	7.67	0.54	7.63	0.54	0.93	0.93	0.92	0.33	0.74

*Structural MRI data was available for one of the participants who provided no usable fMRI data.

Test-retest reliability, anatomically defined ROIs

Reliabilities obtained for maximum values within the anatomical ROIs were comparable to the test-retest reliabilities established for maximum values within the functional ROIs (see Table S4.2). Reliabilities for the contrast familiar vs. unfamiliar were poor for the whole sample, and for left-handed and right-handed participants separately, in all ROIs (FFA, IFG, STG; -.33 ≤ ICC ≤ .35, with the exception of FFA activity for left-handed participants ICC= .44). FFA activity related to face processing (contrast: face vs. scrambled) was fair to good both across the entire sample and for left- and right handed participants separately (.41 ≤ ICC ≤ .62). For V1, reliability was poor for all contrasts vs. fixation cross for the entire sample (.24 ≤ ICC ≤.36) and for right-handed participants (.08 ≤ ICC ≤ .20). For left-handed participants reliability was fair (.40 ≤ ICC ≤ .56; with the exception of NeutralUnfamiliar vs. fixation cross: ICC = .29). Similar to the results obtained for maximum values within the functional masks, increasing the number of trials did not clearly increase reliability values, and even seemed associated with decreasing reliabilities for

V1 (note the usually fair reliabilities for 78 trials, .39 ≤ ICC ≤ 51). In general, test-retest reliabilities for left-handed participants were slightly higher than for right-handed participants, but differences were not significant.

Table S4.2. Test-retest reliabilities for the anatomical masks with maximum values for the whole sample for the first third (78 trials N= 42[1]), the first two thirds of the task (156 trials N= 42[1]), and the complete task (234 trials N = 41), and for left- (N= 20) and right-handed (N= 21) participants separately.

ROI	Contrast	Number of trials			Left-handed	Right-handed	Fisher's r to z	
		78	156	234			Z	p
V1	ThreatFamiliar vs. fix	0.39	0.36	0.36	0.56	0.12	1.49	0.14
	ThreatUnfamiliar vs. fix	0.40	0.39	0.34	0.48	0.14	1.12	0.26
	ThreatScrambled vs. fix	0.51	0.37	0.34	0.51	0.14	1.24	0.22
	NeutralFamiliar vs. fix	0.47	0.47	0.30	0.47	0.08	1.29	0.20
	NeutralUnfamiliar vs. fix	0.44	0.40	0.24	0.29	0.20	0.27	0.79
	NeutralScrambled vs. fix	0.44	0.28	0.27	0.40	0.11	0.96	0.34
FFA	Familiar vs. Unfamiliar	0.27	0.37	0.35	0.44	0.23	0.70	0.48
	Face vs. Scrambled	0.50	0.58	0.54	0.41	0.62	-0.82	0.41
IFG	Familiar vs. Unfamiliar	0.16	0.03	-0.25	-0.22	-0.30	0.24	0.81
STG	Familiar vs. Unfamiliar	0.06	-0.11	-0.10	-0.05	-0.33	0.87	0.38

[1] For one participant, data was only available for the first and the second part of the task, since this participant fell asleep during the third part.
fix = fixation cross

Test-retest reliability, mean and median cope values

As shown in Tables S4.3 and S4.4, test-retest reliability for mean and median values within the functionally defined ROIs were lower than reliability scores obtained for maximum values. Good test-retest reliabilities were obtained only for activity related to face processing within the FFA (.54 ≤ ICC ≤ .72 [contrast face vs. scrambled]). All other ROIs and contrasts showed poor reliability (ICCs ≤ .39, with three exceptions among left-handed participants: ICC = .54 [median, IFG] ICC = .51 [mean, IFG], ICC = .43 [mean, FFA, familiar vs. unfamiliar]) Comparable to the results for maximum values, ICCs did not consistently increase with increasing numbers of trials and ICCs obtained for 78, 156, and 234 trials differed only slightly from each other (see Tables S4.3 and S4.4). Again, test-retest reliabilities for left-handed participants were higher than for right-handed participants, but after correcting for multiple testing, differences were not significant.

Table S4.3 Test-retest reliabilities for the functional masks with mean values for the whole sample for the first third (78 trials N= 42[1]), the first two thirds of the task (156 trials N= 42[1]), and the complete task (234 trials N = 41), and for left- (N= 20) and right-handed (N= 21) participants separately.

ROI	Contrast	Number of trials			Left-handed	Right-handed	Fisher's r to z	
		78	156	234			Z	p
V1	ThreatFamiliar vs. fix	0.11	0.09	0.05	0.11	-0.01	0.36	0.72
	ThreatUnfamiliar vs. fix	0.12	0.05	0.09	0.17	0.00	0.53	0.60
	ThreatScrambled vs. fix	0.28	0.10	0.09	0.16	0.02	0.42	0.68
	NeutralFamiliar vs. fix	0.11	0.09	0.07	0.10	0.04	0.16	0.87
	NeutralUnfamiliar vs. fix	0.14	0.14	0.11	0.20	0.04	0.48	0.63
	NeutralScrambled vs. fix	0.13	0.10	0.07	0.12	0.02	0.29	0.77
FFA	Familiar vs. Unfamiliar	0.19	0.23	0.21	0.43	-0.09	1.60	0.11
	Face vs. Scrambled	0.54	0.61	0.65	0.56	0.71	-0.76	0.45
IFG	Familiar vs. Unfamiliar	0.07	0.07	0.02	0.51	-0.23	2.34	0.02*
STG	Familiar vs. Unfamiliar	-0.02	0.15	0.23	0.27	0.16	0.35	0.73

[1] For one participant, data was only available for the first and the second part of the task, since this participant fell asleep during the third part.

*Difference was not significant after applying the Benjamini-Hochberg procedure to correct for multiple testing.

fix = fixation cross

Table S4.4 Test-retest reliabilities for the functional masks with median values of the whole sample for the first third (78 trials N= 42[1]), the first two thirds of the task (156 trials N= 42[1]), and the complete task (234 trials N = 41), and for left- (N= 20) and right-handed (N= 21) participants separately.

ROI	Contrast	Number of trials			Left-handed	Right-handed	Fisher's r to z	
		78	156	234			Z	p
V1	ThreatFamiliar vs. fix	0.11	0.10	0.04	0.10	0.00	0.32	0.75
	ThreatUnfamiliar vs. fix	0.10	0.04	0.08	0.17	0.00	0.50	0.62
	ThreatScrambled vs. fix	0.22	0.07	0.05	0.10	0.00	0.30	0.76
	NeutralFamiliar vs. fix	0.07	0.09	0.07	0.08	0.06	0.05	0.96
	NeutralUnfamiliar vs. fix	0.10	0.14	0.13	0.23	0.05	0.53	0.60
	NeutralScrambled vs. fix	0.05	0.09	0.05	0.07	0.03	0.10	0.92
FFA	Familiar vs. Unfamiliar	0.23	0.20	0.17	0.39	-0.11	1.53	0.13
	Face vs. Scrambled	0.60	0.63	0.66	0.57	0.72	-0.80	0.42
IFG	Familiar vs. Unfamiliar	0.10	0.06	0.04	0.54	-0.22	2.42	0.02*
STG	Familiar vs. Unfamiliar	-0.01	0.10	0.25	0.28	0.17	0.34	0.73

[1] For one participant, data was only available for the first and the second part of the task, since this participant fell asleep during the third part.

*Difference was not significant after applying the Benjamini-Hochberg procedure to correct for multiple testing.

fix = fixation cross

Test-retest reliability for consistency

Table S4.5 displays test-retest reliabilities for consistency, calculated for maximum cope values within functional ROIs. ICCs for consistency were generally comparable to ICCs for absolute agreement (see Table S4.5). For the contrast familiar vs. unfamiliar (FFA, IFG, STG), ICCs for consistency were poor (-.16 ≤ ICC ≤ .34), with the exception of fair reliability for FFA activity obtained for left-handed participants (ICC = .53). FFA activity related to face processing (contrast: face vs. scrambled) showed good test-retest reliability (.65 ≤ ICC ≤ .74). For V1, ICCs for right-handed participants were poor (.02 ≤ ICC ≥ .18), but ICCs for left-handed participants were fair to excellent (.52 ≤ ICC ≤ .84) and thus somewhat higher than ICCs obtained for absolute agreement. Across the entire sample, ICCs were poor to fair for V1 activity (.28 ≤ ICC ≥ .53). ICCs were generally higher for left-handed than for right-handed participants, with significant differences obtained in V1 for the contrasts ThreatFamilar vs. fixation cross, ThreatScrambled vs. fixation cross, and NeutralScrambled vs. fixation cross after correcting for multiple testing.

Table S4.5. Test-retest reliabilities for maximum values of the whole sample for the first third (78 trials N= 42[1]), the first two thirds of the task (156 trials N= 42[1]), and the complete task (234 trials N = 41), and for left- (N= 20) and right-handed (N= 21) participants separately (234 trials).

ROI	Contrast	Number of trials			Left-handed	Right-handed	Fisher's r to z	
		78	156	234			Z	p
V1	ThreatFamiliar vs. fix	0.44	0.41	0.43	0.77	0.07	2.81	0.01**
	ThreatUnfamiliar vs. fix	0.41	0.43	0.38	0.64	0.06	2.06	0.04*
	ThreatScrambled vs. fix	0.55	0.49	0.53	0.84	0.18	3.07	0.00**
	NeutralFamiliar vs. fix	0.48	0.50	0.34	0.65	0.02	2.23	0.03*
	NeutralUnfamiliar vs. fix	0.45	0.46	0.28	0.52	0.18	1.17	0.24
	NeutralScrambled vs. fix	0.49	0.38	0.44	0.82	0.10	3.12	0.00**
FFA	Familiar vs. Unfamiliar	0.13	0.25	0.29	0.53	0.04	1.63	0.10
	Face vs. Scrambled	0.65	0.65	0.71	0.66	0.74	-0.47	0.64
IFG	Familiar vs. Unfamiliar	0.06	0.11	0.01	0.27	-0.16	1.3	0.19
STG	Familiar vs. Unfamiliar	0.12	0.08	0.18	0.34	-0.04	1.17	0.24

[1] For one participant, data was only available for the first and the second part of the task, since this participant fell asleep during the third part.

fix = fixation cross

*Difference was not significant after applying the Benjamini-Hochberg procedure to correct for multiple testing.

**Difference was significant after applying the Benjamini-Hochberg procedure to correct for multiple testing.

Within-session reliability

Table S4.6 presents within-session reliabilities for maximum values (functional ROIs). In session 1, reliability of V1 activity was fair to excellent across the entire sample (.49 ≤ ICC ≤ .77) and for left- and right-handed participants separately (.47 ≤ ICC ≤ .80). In session 2, reliability of V1 activity was fair to excellent for left-handed participants (.41 ≤ ICC ≤ .77), but poor for right-handed participants (ICC ≤ .31, except NeutralScrambled vs. fixation: ICC

= .46), resulting in poor to good reliabilities across the entire sample (.34 ≤ ICC ≤ .61). Reliability of FFA activity related to face processing (contrast face vs. scrambled) was clearly higher in session 1 (.51 ≤ ICC ≤ .67; fair to good) compared to session 2 (.24 ≤ ICC ≤ .49; poor to fair). For the contrast familiar vs. unfamiliar, we obtained poor reliability values in both sessions for all ROIs (session 1: .01 ≤ ICC ≤ .34; except for IFG activity in right-handed participants, ICC= .42; session 2: -.14 ≤ ICC ≤ .25). With respect to handedness, we did not obtain systematic differences in ICC values between left-handed and right-handed participants in session 1. However, in session 2 reliabilities were systematically higher for left-handed than for right-handed participants, although differences were significant only for V1 activity for the contrast NeutralUnfamiliar vs. fixation cross ($p < .01$) after correcting for multiple testing using the Benjamini-Hochberg procedure.

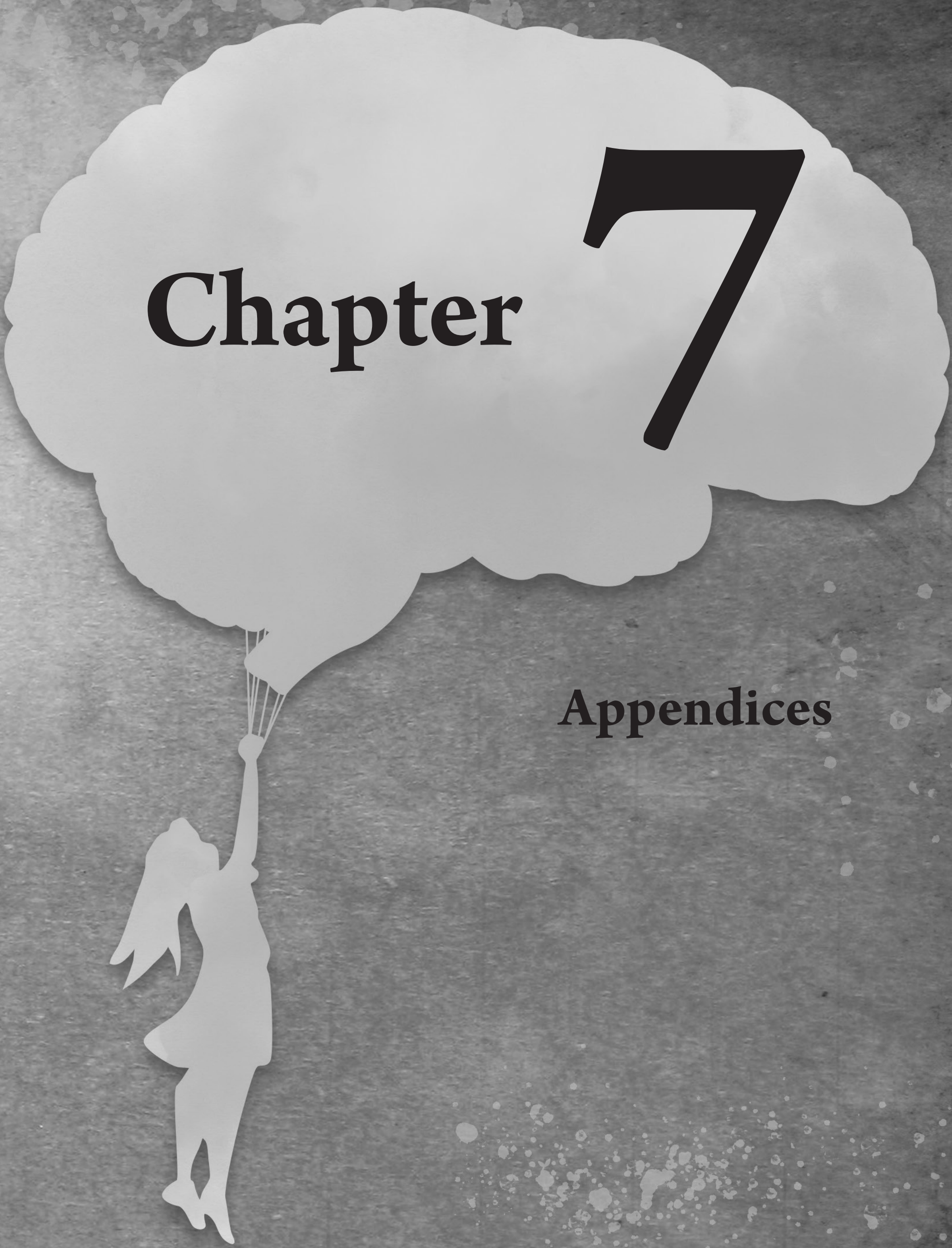
Table S4.6. Test-retest reliabilities for the functional masks with max values for the whole group, and for left- (N = 20) and right-handed (N= 21) participants separately within session 1 and within session 2.

ROI	Contrast	Session 1			Session 2		
		Whole sample	Left-handed	Right-handed	Whole sample	Left-handed	Right-handed
V1	ThreatFamiliar vs. fix	0.65	0.64	0.66	0.34	0.44	0.24
	ThreatUnfamiliar vs. fix	0.49	0.52	0.48	0.45	0.73*	0.15
	ThreatScrambled vs. fix	0.69	0.64	0.74	0.52	0.73	0.31
	NeutralFamiliar vs. fix	0.67	0.77	0.58	0.35	0.41	0.28
	NeutralUnfamiliar vs. fix	0.61	0.47	0.77	0.38	0.77**	0.13
	NeutralScrambled vs. fix	0.77	0.72	0.80	0.61	0.77	0.46
FFA	Familiar vs. Unfamiliar	0.16	0.11	0.22	0.08	0.13	0.02
	Face vs. Scrambled	0.60	0.51	0.67	0.36	0.24	0.49
IFG	Familiar vs. Unfamiliar	0.34	0.17	0.42	0.00	0.13	-0.12
STG	Familiar vs. Unfamiliar	0.17	0.01	0.28	0.13	0.25	-0.14

Effects for Handedness are calculated with Fisher's r to z transformation.

*Difference was not significant after applying the Benjamini-Hochberg procedure to correct for multiple testing.

**Difference was significant after applying the Benjamini-Hochberg procedure to correct for multiple testing.



Chapter 7

Appendices

Nederlandse samenvatting (Summary in Dutch)

Mensen zijn sociale dieren. Het is daarom niet verwonderlijk dat onze sociale interacties een grote invloed hebben op ons geluk en welbevinden (Helliwell, 2003; Helliwell & Putnam, 2004; Pichler, 2006). De effectiviteit van onze interacties met anderen kan door verschillende factoren, zoals onze verbale en non-verbale communicatievaardigheden en onze vaardigheid om gezichten van anderen accuraat te verwerken, worden beïnvloed. De accurate verwerking van gezichten stelt ons in staat om bijvoorbeeld het geslacht en de emotionele toestand van een ander te bepalen en het helpt bij het herkennen van bekenden. Hierdoor kunnen wij ons gedrag op anderen afstemmen. Daarnaast helpt het accuraat waarnemen van gezichten ook bij het herkennen van onze verwanten. Als de gezichten van twee individuen sterk op elkaar lijken impliceert dit namelijk nauwe genetische verwantschap tussen deze twee personen (Alvergne , Faurie, & Raymond, 2010; Bressan & Grassi, 2004; Kaminski, Dridi, Graff, & Gentaz, 2009; Liebermann, 2007; Maloney & Dal Martello, 2006). Hoeveel wij op iemand anders lijken kan ook beïnvloeden hoe wij op diegene reageren. In eerder onderzoek waren volwassen proefpersonen bijvoorbeeld sneller bereid om geld aan een kind te besteden of tijd met een kind door te brengen naarmate dat meer op hen leek (Bressan, Bertamini, Nalli, & Zanutto, 2009; DeBruine, 2004; Platek et al., 2004).

Hoe mensen gezichten verwerken en op gezichten reageren verschilt per individu. Veel factoren, zoals iemands temperament, de context waarin een gezicht wordt waargenomen en jeugdervaringen, kunnen beïnvloeden hoe wij op gezichten reageren. Op welke manier deze factoren de reacties van individuen op gezichten beïnvloeden is echter nog niet goed onderzocht. In dit proefschrift wordt daarom onderzocht hoe sommige van deze factoren – vroege negatieve ervaringen met opvoeders, individuele verschillen in de mate van empathie, en een bedreigende context – reacties op kindergezichten beïnvloeden. In dit proefschrift is gebruik gemaakt van kindergezichten die in verschillende mate op de proefpersonen leken.

Dit proefschrift

Voor het onderzoek beschreven in dit proefschrift zijn foto's van kinderen zo bewerkt dat deze gingen lijken op de gezichten van de proefpersonen. Op deze manier is genetische verwantschap gesuggereerd. Hiervoor werd een foto van de proefpersoon samengevoegd met een foto van een onbekend kind. Op deze manier ontstond er voor elke proefpersoon een foto van een kindergezicht dat op het gezicht van deze proefpersoon leek. Het kindergezicht vertoonde 50% gelijkenis met het gezicht van de proefpersoon. Het doel was om door de gelijkenis met het kindergezicht ‘ouderlijke’ reacties bij de proefpersonen op te wekken. Om te onderzoeken welke processen in de hersenen een rol spelen bij ouderlijke reacties is in dit proefschrift gebruik gemaakt van functionele magnetische kernspintomografie (*functional magnetic resonance imaging*: fMRI).

Met fMRI kan hersenactiviteit gemeten worden. Hersengebieden die betrokken zijn bij het verwerken van bepaalde stimuli, zoals gezichten, laten in reactie op deze stimuli een verhoogde activiteit zien. Om hersenactiviteit met fMRI te bestuderen zijn in onze studie kindergezichten op een scherm getoond terwijl de proefpersonen in een MRI-scanner lagen. Naast het kindergezicht dat op het gezicht van de proefpersoon leek werd tijdens de fMRI-taak ook een kindergezicht getoond dat niet op de proefpersoon leek. Op deze wijze was het mogelijk om verschillen in neurale reacties op kindergezichten die niet op de proefpersonen leken en kindergezichten die wèl op de proefpersonen leken te onderzoeken. Om de unieke reacties van de hersenen in reactie op een bepaalde stimulus te kunnen bepalen wordt in fMRI onderzoek de mate van hersenactiviteit in verschillende condities met elkaar vergeleken (in onze studie: kinderen die niet op de proefpersonen leken vs. kinderen die wèl op de proefpersonen leken).

De kindergezichten zijn in de fMRI-taak met bedreigende situaties gecombineerd, omdat we verwachtten dat die ouderlijke beschermende reacties oproepen (George & Salomon, 2009). De uitkomsten van eerdere studies wijzen erop dat bedreigende stimuli mogelijkerwijs zonder bewuste waarneming van deze stimuli door de hersenen verwerkt worden (Almeida, Pajtas, Mahon, Nakayama, & Caramazza, 2013; Morris, Öhman & Dolan, 1998; Whalen et al., 1998).

Om dit te onderzoeken werd in de huidige studie voorafgaand aan een kindergezicht telkens een afbeelding van een bedreigende of een neutrale situatie op het scherm getoond. Neutrale scènes werden getoond om hersenactiviteit in reactie op bedreigende situaties te kunnen vergelijken met hersenactiviteit in reactie op neutrale situaties. De proefpersonen konden de bedreigende en neutrale scènes niet bewust waarnemen. De scènes werden namelijk erg kort op het scherm gepresenteerd en gemaskeerd door een plaatje met een felgekleurd, cirkelvormig patroon dat onmiddellijk voor en na de bedreigende en neutrale scènes op het scherm te zien was. Op deze manier was het mogelijk om te onderzoeken of bedreigende stimuli verwerkingsprocessen in het brein kunnen beïnvloeden zonder bewuste waarneming van deze stimuli. Alle bedreigende scènes toonden een persoon die door iemand anders werd aangevallen, bijvoorbeeld een persoon die door iemand anders de keel werd toegeknepen. Als neutrale scènes waren niet-emotionele situaties afgebeeld, zoals twee personen die met elkaar in gesprek waren.

De proefpersonen voerden naast de fMRI-taak ook een taak achter de computer uit waarin zij kindergezichten die verschilden in de mate waarop zij op de proefpersonen leken beoordeelden op een aantal positieve en negatieve criteria, zoals aantrekkelijkheid en onbetrouwbaarheid. Tijdens de computertaak werden dezelfde kindergezichten getoond als tijdens de fMRI-taak, plus een kindergezicht dat nog meer op de proefpersoon leek. Hierdoor ontstonden drie condities: geen gelijkenis, 50% gelijkenis met het gezicht van de proefpersoon, en 75% gelijkenis met het gezicht van de proefpersoon. Op deze manier is onderzocht of kindergezichten die meer op de proefpersonen leken positiever en minder negatief werden beoordeeld dan kindergezichten die minder op de proefpersonen leken. Daarnaast is inzicht verkregen in de samenhang tussen hersenactiviteit in reactie op de kindergezichten (gemeten tijdens de fMRI-taak) en beoordelingen van deze gezichten op een aantal negatieve en positieve criteria (gemeten tijdens de computertaak).

Gezichtsverwerking en gelijkenis

Gezichtsverwerking is een complex proces. Het is daarom niet verwonderlijk dat een uitgebreid netwerk van hersengebieden betrokken is bij het verwerken van en reageren op gezichten. De inferieure occipitale gyrus, een deel van de occipitaalkwab (een gebied aan de achterkant van het brein), is betrokken bij het verwerken van de fysieke kenmerken van een gezicht. De superieure temporale gyrus (STG), die deel uitmaakt van de temporaalkwab (een gebied in het midden van het brein, net achter de oren), speelt een rol bij het verwerken van veranderlijke kenmerken van gezichten, zoals blikrichting en lipbewegingen. De *fusiform face area* (FFA) wordt geassocieerd met het verwerken van onveranderlijke kenmerken van een gezicht om de identiteit van een persoon te bepalen. De FFA ligt in de fusiforme gyrus, die deels in de occipitaalkwab en deels in de temporaalkwab van het brein ligt (Andrews & Ewbank, 2004; Andrews & Schluppeck, 2004; Grill-Spector, Knouf & Kanwisher, 2006; Gobbini & Haxby, 2007; Liu, Harris, & Kanwisher, 2010; Nichols, Betts, & Wilson, 2010; Pitcher, Walsh, Yovel, & Duchaine, 2007; Yovel & Kanwisher, 2005). Daarnaast zijn ook andere hersengebieden belangrijk voor het accuraat verwerken van gezichten, zoals de amygdala bij het verwerken van emotionele informatie (Gobbini, Leibenluft, Santiago & Haxby, 2004).

Hoofdstuk 2 van dit proefschrift beschrijft een studie waarin bij 45 jongvolwassen vrouwen zonder eigen kinderen veranderingen in hersenactiviteit in reactie op kindergezichten die op de proefpersonen leken onderzocht zijn. In reactie op gezichten die op de proefpersonen leken was een verhoogde activiteit in de FFA, de infero-laterale occipitale gyrus en de occipitale pool te zien, hersengebieden die betrokken zijn bij de visuele verwerking van stimuli, en meer specifiek bij gezichtsverwerking. Gezichten die op de proefpersonen leken werden dus intensiever verwerkt. Dit suggereert een preferentiële verwerking van de gezichten die lijken op de proefpersonen (Natu & O'Toole, 2011). Daarnaast beoordeelden proefpersonen de kindergezichten die meer op hen leken ook minder negatief. Mogelijkerwijs leiden de minder negatieve gevoelens die individuen ervaren bij het zien van gezichten die meer op henzelf lijken tot betere zorg voor verwante individuen. Verwanten lijken namelijk gemiddeld meer op elkaar dan individuen die niet aan elkaar verwant zijn. Minder negatieve gevoelens bij het zien van gezichten die op jezelf lijken zouden tot betere zorg voor biologische kinderen vergeleken met niet-biologische kinderen kunnen leiden (Alvergne et al., 2007; Bressan & Grassi, 2004; DeBruine et al., 2009; Maloney & DalMartello, 2006). Daarnaast was in reactie op kindergezichten die op de proefpersonen leken een verhoogde activiteit in de rechter inferieure frontale gyrus (IFG), mediale frontale gyrus (MFG) en insula te zien. De IFG en de MFG zijn betrokken bij het verwerken van zelf-gerelateerde informatie (d.w.z. informatie over jezelf, met inbegrip van uiterlijke en innerlijke kenmerken). Daarnaast zijn de IFG en insula betrokken bij de regulatie van emoties (Devue & Bredart, 2011). Deze hersengebieden spelen ook een belangrijke rol bij ouderlijk gedrag (Swain et al., 2014).

De STG, het planum temporale en de posterieure supramarginale gyrus evenals de superieure parietaalkwab, cuneus en precuneus, lieten daartegen een verminderde activiteit zien in reactie op kindergezichten die op de proefpersoon leken. Deze hersengebieden maken deel uit van het socio-emotionele netwerk van het brein en zijn gerelateerd aan *Theory of Mind* (in het bijzonder de STG). Onder Theory of Mind wordt het vermogen verstaan om mentale toestanden, zoals overtuigingen, wensen en intenties, aan anderen toe te schrijven en zich daarin te kunnen verplaatsen (Frith & Frith, 1999; Schurz, Radua, Aichhorn, Richlan, & Perner, 2014). Een mogelijke verklaring voor de verminderde activiteit in reactie op kindergezichten die op de proefpersonen leken, is dat individuen er wellicht automatisch van uitgaan dat mensen die uiterlijke overeenkomsten met hen vertonen ook innerlijk op hen lijken, en soortgelijke wensen, intenties en overtuigingen hebben. Als gevolg daarvan kunnen mensen minder noodzaak voelen om de mentale toestanden van individuen die op hen lijken grondig te onderzoeken, wat de verminderde activiteit in Theory of Mind gebieden in reactie op kindergezichten die op de proefpersonen leken kan verklaren. In overeenstemming met deze verklaring is er in een eerdere studie verminderde activiteit in de superieure temporale sulcus en een aantal andere hersengebieden gevonden als moeders foto's van hun eigen kinderen zagen in vergelijking met foto's van bekende, maar niet verwante kinderen (Bartels & Zeki, 2004). De noodzaak om de mentale toestanden te bepalen van andere individuen wordt blijkbaar minder gevoeld als die individuen erg bekend (of nauw verwant) zijn.

In Hoofdstuk 3 is bij 41 jongvolwassen vrouwen de invloed van de mate van gelijkenis van kindergezichten met de proefpersonen op positieve en negatieve beoordelingen van deze gezichten bestudeerd. Daarnaast is de samenhang tussen beoordelingen van de kindergezichten en hersenactiviteit in reactie op de gezichten onderzocht. Er was een verband tussen de intensiteit waarmee proefpersonen de kindergezichten verwerkten en negatieve beoordelingen van kindergezichten. Bij proefpersonen die meer *love withdrawal* in hun kindertijd hadden ervaren was een sterkere activiteit in de FFA gerelateerd aan minder negatieve beoordelingen van de kindergezichten die meer op de proefpersoon leken. De effecten van gelijkenis waren dus sterker voor proefpersonen die de kindergezichten intensiever verwerkten (te zien aan een hogere activiteit in de FFA). Hogere activiteit in de FFA impliceert dus mogelijk een grotere gevoeligheid voor gelijkenis.

Individuele verschillen en context

Als een kind in gevaar is of veel stress ervaart, hebben ouders een sterke behoefte om het kind te beschermen en voor het kind te zorgen (George & Solomon, 2008). Mogelijkerwijs worden beschermend gedrag en de neurobiologische processen die aan het beschermend gedrag ten grondslag liggen ook opgewekt als een kind dat op de proefpersoon lijkt in gevaar is of bedreigd wordt. In Hoofdstuk 2 is onderzocht of subliminaal (d.w.z. te kort om bewust te worden waargenomen) gepresenteerde afbeeldingen van bedreigende situaties die onmiddellijk voor de afbeelding van een kind dat op de proefpersoon lijkt op het scherm verschenen tot veranderingen in hersenactiviteit zouden leiden, zoals een verhoogde activiteit in de amygdala. Er zijn echter geen meetbare verschillen gevonden tussen reacties op bedreigende en neutrale situaties.

Er zijn verschillende verklaringen mogelijk voor het ontbreken van significante effecten van bedreigende situaties die in combinatie met de kindergezichten werden gepresenteerd op de hersenactiviteit. Ten eerste waren de neutrale en bedreigende situaties die in dit proefschrift zijn gebruikt vrij complex (op de foto's waren meerdere objecten en personen te zien) wat de onbewuste identificatie van de bedreigende en de neutrale lading van de scènes mogelijkerwijs heeft belemmerd. Een andere verklaring is dat de bedreigende scènes geen beschermende reacties in de proefpersonen hebben uitgelokt. Er zijn echter ook studies die in twijfel trekken of er wel een systeem in het brein bestaat dat bedreigende stimuli automatisch en zonder bewuste waarneming kan verwerken (Pessoa & Adolphs, 2010; Hoffmann et al., 2012). In studies waarin wel meetbare effecten van deze stimuli op gedrag en hersenactiviteit zijn gevonden (Pessoa, Japee & Ungerleider, 2005) werden de bedreigende stimuli mogelijkerwijs niet echt subliminaal gepresenteerd. Er is echter nog meer onderzoek nodig naar de mogelijke effecten van subliminaal gepresenteerde stimuli op veranderingen in hersenactiviteit.

De reacties van individuen op een kind in een bedreigende context kunnen ook door hun mate van empathie, oftewel het vermogen om mee te leven met anderen, worden beïnvloed. In Hoofdstuk 2 van dit proefschrift is daarom de mate van empathie van de proefpersonen met een vragenlijst in kaart gebracht. Individuen met een hogere mate van empathie reageren mogelijkerwijs sterker op een kind dat veel stress ervaart of in potentieel gevaar is. Er werden echter geen significante effecten van empathie op hersenactiviteit gevonden. Gezien in het huidige onderzoek ook geen meetbare veranderingen in hersenactiviteit in reactie op de bedreigende situaties zijn gevonden, is het ontbreken van significante effecten van empathie op hersenactiviteit niet verwonderlijk. De resultaten impliceren namelijk dat de proefpersonen de bedreigende context waarschijnlijk niet hebben waargenomen.

Reacties op kinderen die op ons lijken zouden bovendien door onze vroege jeugdervaringen met opvoeders beïnvloed kunnen worden. Vroege ervaringen met opvoeders vormen onze ideeën over relaties met anderen en beïnvloeden hoe wij op voor ons belangrijke anderen, zoals familieleden, reageren (Mikulincer, Shaver, Gilath, & Nitzberg, 2005). Sommige opvoedstrategieën, zoals psychologische controle en love withdrawal, lijken bovendien negatieve effecten op het latere welbevinden en functioneren van kinderen te hebben. Bij het gebruik van love withdrawal in de opvoeding hangt de liefde en affectie van de ouder af van het succes en gedrag van het kind. Er ontstaat hierdoor een verband tussen gehoorzaamheid en prestaties van het kind enerzijds en zijn of haar relatie met de ouder anderzijds. Het regelmatige gebruik van love withdrawal hangt samen met symptomen van angst en depressie en een lagere zelfcontrole bij kinderen en adolescenten, en faalangst bij volwassenen (Elliot & Thrash, 2004; Hill & Bush, 2001; Mandara & Pikes, 2008). Daarnaast zijn er ook veranderingen in de verwerking van en reacties op emotionele gezichten bij jongvolwassenen gevonden die love withdrawal hebben ervaren (Huffmeijer, Tops, Alink, Bakermans-Kranenburg, & Van IJzendoorn, 2011, Huffmeijer et al. 2013).

In Hoofdstuk 2 zijn de relaties tussen ervaringen van love withdrawal en hersenactiviteit in reactie op kindergezichten die op de proefpersonen leken onderzocht. Ervaringen van love withdrawal zijn met een vragenlijst gemeten. In Hoofdstuk 3 is de samenhang tussen hersenactiviteit, ervaringen van love withdrawal, en negatieve en positieve beoordelingen van kindergezichten die verschilden in de mate waarop zij op de proefpersonen leken bestudeerd. Proefpersonen vulden deze vragenlijst aan het begin van het onderzoek over hun vader en hun moeder in.

De verschillen in hersenactiviteit in de rechter infero-laterale occipitaalkwab, de FFA en de occipitale pool in reactie op kindergezichten die op de proefpersonen leken en kindergezichten die niet op de proefpersonen leken waren groter naarmate proefpersonen meer love withdrawal hadden ervaren. Daarnaast was er met toenemende gelijkenis van de kindergezichten met de proefpersonen een sterkere daling in negatieve beoordelingen te zien bij proefpersonen die gezichten intensiever verwerkten (te zien aan een hogere activiteit in de FFA) en meer love withdrawal hadden ervaren. De effecten van hoeveel de kindergezichten op de proefpersoon leken op de beoordelingen van die gezichten waren dus meer uitgesproken in individuen die meer love withdrawal hadden ervaren. De resultaten komen ook overeen met de uitkomsten uit eerdere elektro-encefalogram (EEG) studies, waarin proefpersonen die meer love withdrawal hebben ervaren emotionele gezichten intensiever verwerkten (Huffmeijer et al., 2011, 2013). De uitkomsten uit eerdere studies en dit proefschrift duiden er dus op dat ervaringen van love withdrawal de verwerking van sociaal-relevante stimuli, zoals emotionele gezichten en de gezichten van 'familieleden', kunnen beïnvloeden.

Test-hertest betrouwbaarheid

Om conclusies uit de resultaten van fMRI-onderzoeken te kunnen trekken moeten de verkregen data betrouwbaar zijn, wat betekent dat de metingen nauwkeurig en herhaalbaar zijn. Een onderdeel van betrouwbaarheid is dus de *test-hertest* betrouwbaarheid, waarbij onderzocht wordt of een meting die na een bepaalde periode opnieuw wordt gedaan, dezelfde resultaten geeft (Feldt & Brennan, 1989). Omdat maar weinig studies de test-hertest betrouwbaarheid van fMRI-data in reactie op gezichten hebben onderzocht, is in de studie beschreven in Hoofdstuk 4 de betrouwbaarheid van de fMRI-data uit dit proefschrift bestudeerd. De betrouwbaarheid van veranderingen in hersenactiviteit in de IFG, de STG en de FFA werd gedurende verschillende informatieverwerkingsprocessen (herkennen van gezichten, verwerking van gelijkenis) onderzocht.

Het herkennen van gezichten gebeurt in een vroege verwerkingsfase (Eimer, 2000), die sterk gelokaliseerd is in de FFA (Gauthier, Skudlarski, Gore, & Anderson, 2000; Haxby et al., 2001; Gobbini & Haxby, 2007). Veranderingen in hersenactiviteit in de FFA geassocieerd met het verwerken van gezichten lieten een stabiele test-hertest betrouwbaarheid zien. De bekendheid van een gezicht beïnvloedt latere verwerkingsfases, en daarbij zijn verschillende hersengebieden betrokken die een rol spelen bij diverse cognitieve processen (Natu & O'Toole, 2011). De test-hertest betrouwbaarheid van veranderingen in hersenactiviteit geassocieerd met de bekendheid van gezichten was echter gering in alle onderzochte hersengebieden (FFA, IFG, STG). Veranderingen in hersenactiviteit in vroege verwerkingsfases, zoals de identificatie van een gezicht, kunnen dus mogelijkerwijs betrouwbaarder gemeten worden dan veranderingen die optreden tijdens latere verwerkingsfases die gerelateerd zijn aan subtielere verschillen tussen stimuli, zoals de mate van gelijkenis en bekendheid van een gezicht.

Omdat linkshandigen vaak worden uitgesloten van fMRI-onderzoek terwijl 10% van de mensen linkshandig is (Willems, Van der Haegen, Fisher, & Francks, 2014; McManus, 2009), zijn ook mogelijke verschillen in test-hertest betrouwbaarheid tussen links- en rechtshandige proefpersonen bestudeerd. Verschillen tussen links- en rechtshandigen waren doorgaans niet statistisch significant. Voor fMRI-onderzoek dat gebruik maakt van soortgelijke taken als de taak die in dit proefschrift wordt beschreven zijn zorgen over betrouwbaarheid daarom geen reden om linkshandigen uit te sluiten. Tot slot is onderzocht of het toevoegen van extra datapunten (d.w.z. dat de kindergezichten en de neutrale en bedreigende situaties vaker getoond werden en de taak dus langer werd) tot een hogere betrouwbaarheid zou leiden. Eerder onderzoek heeft namelijk uitgewezen dat het toevoegen van extra datapunten de betrouwbaarheid van fMRI-taken kan verhogen (Bennett & Miller, 2010). Er was echter geen consistente toename in de test-hertest betrouwbaarheid met een toenemend aantal datapunten, wat suggereert dat het aanbieden van meer stimuli in fMRI-onderzoek niet altijd tot hogere betrouwbaarheden leidt. Onze uitkomsten impliceren daarom dat onderzoekers zorgvuldig de voor- en nadelen van het toevoegen van extra datapunten in fMRI-studies moeten afwegen. Lage betrouwbaarheden zijn een veel voorkomend probleem in fMRI-onderzoek (Bennett & Miller, 2010), en het optimaliseren van de betrouwbaarheid van fMRI-data in toekomstig onderzoek is daarom van groot belang.

Vervolgonderzoek

Het doel van de huidige studie was om eerste inzichten in de neurale processen onderliggend aan ouderlijk gedrag, en in het bijzonder beschermende reacties, te verkrijgen. Voor toekomstig onderzoek is het belangrijk om deze neurale processen in moeders te bestuderen. Op deze manier kunnen mogelijke verschillen in de uitkomsten tussen moeders en jonge vrouwen zonder eigen kinderen in kaart gebracht worden. Verder zou het interessant zijn om de effecten van gelijkenis met kindergezichten bij mannen te bestuderen, omdat mannen mogelijkerwijs anders dan vrouwen reageren op kindergezichten die wel of niet op henzelf lijken. In sommige studies zijn bijvoorbeeld grotere effecten van gelijkenis bij mannen gevonden dan bij vrouwen. Zo beoordeelden in eerdere studies alleen mannen kindergezichten die op hen leken als meer aantrekkelijk dan kindergezichten die niet op hen leken (Platek, Burch, Panyavin, Wasserman, & Gallup, 2002; Volk & Quinsey, 2002). Daarnaast zijn er bij mannen grotere effecten van gelijkenis met een kindergezicht gevonden op hypothetische beslissingen, zoals de bereidheid om een kind te adopteren (Platek et al., 2002, 2004; Volk & Quinsey, 2002, 2007). Toekomstig onderzoek kan gericht worden op de vraag of er ook verschillen zijn tussen mannen en vrouwen voor wat betreft hun neurale reacties op gelijkenis. Tot slot is het belangrijk om in kwalitatief hoogwaardig fMRI-onderzoek met voldoende grote steekproeven te investeren om meer inzicht in het functioneren van het menselijke brein te verkrijgen.

Conclusie

Om inzicht in het samenspel tussen hersenen en gedrag te vergroten, is in dit proefschrift de relatie tussen neurobiologische reacties op kindergezichten, evaluaties van kinderen die varieerden in de mate waarop zij op de proefpersonen leken, en ervaringen van love withdrawal onderzocht. Dit proefschrift draagt op deze manier bij aan een beter begrip van hoe activiteit in de hersenen gerelateerd is aan gedrag, gedachten en gevoelens. Door de effecten van vroege ervaringen, zoals love-withdrawal, op de neurale verwerking van sociaal-relevante informatie (zoals de gezichten van familieleden) te bestuderen, kan meer inzicht verkregen worden in de lange-termijn effecten van vroege negatieve en positieve ervaringen op neuraal functioneren. De resultaten van dit proefschrift benadrukken bovendien de essentiële waarde van betrouwbaarheidsonderzoek in fMRI-studies. Betrouwbaarheids- en replicatiestudies zijn nodig om ons vertrouwen in de resultaten van fMRI-onderzoek te vergroten.

Acknowledgments

Finally, the big moment has arrived; my dissertation is finished, and I can still hardly believe that it is over! The final challenge now is to find the right words to express my gratitude to everyone who has supported me during the last four years. First of all, my special thanks goes to the participants of my study. Without your collaboration I would not stand here today to defend my thesis. I also want to express my gratitude to Laura and Sandra - with your humor, commitment, and hard work you made the data collection a true pleasure.

Next, I want to thank my dear colleagues that made my Ph.D. so much easier and more enjoyable. I will miss the good conversations, lunch breaks and joint dinners. I especially want to thank Renate and Katharina, you are great friends, and I am grateful that I could share all my Ph.D. sorrows and joys with you. Katharina, thank you for being my paranymph. Gea and Esther, thank you for all your support, help and distraction. Chloe, your fabulous humor was always a reason for me to look forward to go to work. Thank you for always putting a smile on my face. Else and Roxanna, you were great office mates.

It goes without saying that I am also deeply grateful for the support I received from my friends. You are the best friends in the world, thank you for all the laughter and distraction. I particularly want to thank Gepke. With you, I had a friend at my side that could truly understand all my Ph.D. struggles. I am grateful that I can always count on you; thank you for being my paranymph. Meli and Aylin, thank you for the many years of friendship and the wonderful holidays together. Esther, thank you for always having an open ear for me.

Last but not least, my special thanks goes to my family that I can always rely on. Mum, Dad, and Jürgen, without you I would not be the person I am today; thank you for always believing in me. Sarah, I am grateful for having such a wonderful big sister that always kept an eye on me. Felipe, I can hardly express my gratitude for your unconditional love, support and never-ending patience.

Curriculum Vitae

Esther Heckendorf was born on February 23rd 1989 in Andernach, Germany. In 2008, she graduated from the Werner Heisenberg Gymnasium in Neuwied, Germany. After graduation from high school, Esther started to study Psychology at the University of Groningen in the Netherlands. She completed her Bachelor's degree in 2011 with a *cum laude* grade point average. In the same year, Esther started her Master's degree in Neuropsychology at the University of Groningen. In 2012 she did an internship at the Neurological Rehabilitation Centre Friedehorst in Bremen (Germany). In 2012, she completed her Master's degree that was granted *cum laude*. After graduation from university, Esther started her Ph.D. trajectory at the Centre for Child and Family Studies at Leiden University (The Netherlands) in 2013. During her Ph.D. trajectory, Esther conducted research on the effects of subliminally presented threatening stimuli, empathy, and early negative experiences with caregivers on changes in brain activity in response to child faces. The results of her research project are described in this dissertation.

Publications

Heckendorf, E., Bakermans-Kranenburg, M. J., Van IJzendoorn, M. H., Voorthuis, A., & Huffmeijer, R. (in press). Brain Activity and Love withdrawal Moderate Effects of Suggested Kinship on Negative Appraisals. *Evolutionary Behavioral Sciences.*

Heckendorf, E., Huffmeijer, R., Bakermans-Kranenburg, M. J., & Van IJzendoorn, M. H. (2016). Neural processing of familiar and unfamiliar children's faces: Effects of experienced love withdrawal, but no effects of neutral and threatening priming. *Frontiers in Human Neuroscience, 10, 231.*

Manuscripts submitted for publication:

Heckendorf, E., Bakermans-Kranenburg, M. J., Van IJzendoorn, M. H., & Huffmeijer, R. (2018). Neural responses to children's faces: Test -retest reliability of structural and functional MRI. *Submitted for publication.*