The neurocognitive development of social decision-making

Bos, W. van den

THE NEUROCOGNITIVE DEVELOPMENT OF SOCIAL DECISION-MAKING

The research in this thesis was supported by VIDI grant 452-07-011 (Crone)

ISBN 978-90-9025903-1

Printed by Off Page, Amsterdam

The Neurocognitive Development of Social Decision-Making

PROEFSCHRIFT

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. P.F. van der Heijden,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 12 april 2011
klokke 13:45 uur

DOOR

Wouter van den Bos
geboren te Amsterdam

promotiecommissie:

promotoren:
PROF. DR. EVELINE A. CRONE
PROF. DR. ERIC VAN DIJK
PROF. DR. MICHIEL WESTENBERG

overige leden:
PROF. DR. RONALD DAHL
PROF. DR. MAURITS VAN DER MOLEN
PROF. DR. RICHARD RIDDERINKHOF
DR. ALLEN SANFEY

Contents

1. General Introduction

1.1 The Development of Social Decision-Making: A Neuroscientific Perspective

Humans grow up in highly complex social environments, and most of the decisions they make are in the context of social interactions. Already in infancy a large proportion of time is spent interacting with caretakers and over the course of development social interactions become more prevalent and particularly more complex. Social interactions involve a complex set of skills that support; (1) understanding and predicting the content of other minds, (2) building and maintaining relationships, and (3) taking into account social norms. One of the most salient developmental challenges is therefore to develop the ability to monitor and regulate thoughts and actions for adaptive behavior in social interactions. Indeed, it has been hypothesized that the prolonged period of human development and the relatively large neocortex have evolved in order to allow for more complex forms of social behavior (Wilson, 2000; Dunbar, 1998).

The main aim of this thesis is to investigate the hypothesis that the development of social behavior is related to developmental changes in different, but interacting, brain networks. The thesis will focus on the developmental period between late childhood and young adulthood, because this transitional period involves a process of major social-reorientation (Nelson et al., 2005; Blakemore, 2008). Moreover, the use of recently developed imaging techniques, such as Magnetic Resonance Imaging (MRI), indicated that this process of social re-orientation is paralleled by significant structural and functional brain changes (Blakemore, 2008).

Understanding the emergence of social behavior in adolescence is of importance to society, as this is the critical transition period during which children gradually become independent individuals (Steinberg, 2008). Furthermore, investigating how adolescent changes in social behavior are instantiated in functional brain networks has the potential to enhance our understanding of both (1) social development, and (2) the neural correlates of social behavior in general. First, while evidence indicates that changes in social behavior are co-determined by socio-cultural (Greenfield et al. 2003) and internal factors (e.g. hormonal milieu), both must have an impact on the

function of brain networks in order to alter behavior. Thus, understanding the brain-behavior relation may provide a deeper insight in the mechanisms that underlie developmental changes in social behavior. In addition, knowledge of how brain development relates to developmental changes in brain function may constrain or extend current theories of cognitive development (Mareschal et al., 2007). Second, because there are regional differences in trajectories of brain development, adolescent social development may serve as a natural model for the study of how different brain networks contribute to social decision-making in general.

Before turning to the introduction of the empirical chapters, a broader background for this thesis will be sketched. First, two specific aspects underlying adolescent social decision-making are described in more detail (**section 1.2**). The following section describes how these changes are paralleled by structural brain changes (**section 1.3**). These developmental changes will be discussed in the context of neurodevelopmental models that hypothesize that the relation between interregional changes in brain structure and social behavior is mediated via changes in brain function (Nelson et al., 2005; Blakemore, 2008; Johnson, 2011). The next section discusses the advantage of Game Theoretical paradigms (economic games) to study the development of social behavior and its neural underpinnings (**section 1.4**). Subsequently, the neuroimaging literature on studies of social interactions with adults is reviewed (**section 1.5**). These studies have emphasized the involvement of different neural networks in social behavior. Together, the theoretical accounts of social development (**section 1.2**), the neurodevelopment models (**section 1.3**), and the adult neuroimaging studies (**section 1.5**) will function as essential reference points for understanding and interpreting developmental changes in adolescent brain and behavior. Finally, an outline of the chapters of this thesis will be provided (**section 1.6**).

1.2 Adolescent social cognitive development: perspective-taking and self-regulation

Adolescence is the transitional period between childhood and adulthood which is characterized by a unique set of physical, cognitive, emotional, social and neurological changes (Steinberg, 2005; Casey et al., 2008). The onset of adolescence occurs with the start of puberty and is marked by large changes in hormone levels and associated changes in physical appearance (Dahl & Gunner, 2009). The end of adolescence is less well defined and culturally diverse (Choudhury, 2010). It is generally considered to be the moment when the major physical changes have taken place, and an individual has attained an independent adult role within society (Lerner & Steinberg, 2004). For purposes

of this thesis, adolescence is defined as the age period between approximately 10 and 22 years.

Adolescence is characterized by a major process of social-reorientation; there is an increase in time spent with peers, and there are qualitative changes in peer relations (Hartup & Stevens, 1997). The most notable change in the nature of social interactions is the shift from a competitive to a more prosocial[1] attitude (Eisenberg et al., 1991, 1995; O'Brien & Bierman, 1988; Schaffer, 1996; Van Lange, et al., 1997). These changes in social behavior during adolescence are thought to stabilize between middle and late adolescence (Eisenberg et al., 1991, 1995; Schaffer, 1996). Indeed, a prosocial orientation is often considered a marker of attaining adult maturity that is accepted by both adolescents and adults (Eisenberg et al., 2005). This thesis concerns the developmental changes in adolescent social behavior, particularly changes in prosocial behavior in interactions with peers, and how these relate to developmental changes in brain function. The theoretical perspectives, that form the background for understanding the relation between brain and behavioral development, are inspired by traditional theories of cognitive development. Within this tradition two dominant strands of developmental theories can be identified that suggest that developmental changes in prosocial behavior during adolescence are related to the development of an increased capability for; (1) perspective-taking (Eisenberg et al., 1991, 1995; Kohlberg, 1981; Selman, 1980) and (2) self-regulation[2] (Zimmerman, 2000; Nelson et al., 2005; Steinberg, 2009).

Perspective-taking

Several developmental theories that explain adolescent changes in social interactions in terms of an increased capability for social perspective-taking (Kohlberg, 1981; Eisenberg et al., 1995)[3]. In general, these theories posit that during development, adolescents learn to better understand the perspective of the other and to coordinate between the different perspectives of self, others and society, which in turn may lead to changes prosocial behavior (Martin, Sokol &

[1] Prosocial behavior refers to "voluntary actions that are intended to help or benefit another individual or group of individuals" (Eisenberg and Mussen 1989, p. 3). These behaviors include a broad range of activities such as sharing, reciprocating, helping and abiding social norms.

[2] The processes of perspective-taking and self-regulation are not considered *mutually exclusive* or *collectively exhaustive* in explaining social development. That is, they are not *mutually exclusive* because it is not perspective-taking or self-regulation alone but in most cases the two processes together that will determine social development. Further, they are not *collectively exhaustive* because these two processes are also not the only driving forces in social development, for instance affective development (for a broader overview see; Steinberg, 2009; Ernst et al., 2008).

[3] Perspective-taking is sometimes called 'mentalizing', and comprises the ability to recognize others and evaluate their mental states (intentions, desires and beliefs), feelings, enduring dispositions and actions (Blakemore, 2008).

Elfers, 2008). In support of this hypothesis there is experimental evidence that adolescents become more skilled in taking the perspective of others (Choudhury et al., 2006; Dumontheil et al., 2010). These studies show that while the most basic theory of mind tasks are passed at around age four (Frith & Frith, 2007), the ability to take the perspective of the other still develops until late adolescence. More importantly, there is evidence for a modest positive correlation between perspective-taking and prosocial behavior in adolescence (Underwood & Moore, 1987).

Note, however, that an increased perspective-taking ability can also be used for strategic or anti-social purposes, such as lying and cheating (Rotenberg, 1991; Beate & Frith, 1992). Thus, although perspective-taking has generally been related to increases in prosocial behavior in the context of everyday scenarios (Underwood & Moore, 1987) it can, in specific situations, also lead to a decrease in prosocial behavior (e.g. in interactions with disliked peers)

Self-regulation

Self-regulation in context of social behavior refers to the capacity to alter one's own behavior, in accordance to certain standards, ideals or goals either stemming from internal or societal expectations (e.g. personal or social norms; Baumeister & Vohs, 2007). The two important aspects of self-regulation are *monitoring* and *adaptation.* First, monitoring is necessary because the social environment is dynamic and constantly changing over the course of performance (Zimmerman, 2000). An important aspect of monitoring behavior is attending to, and processing, internally or externally generated feedback signals. Second, feedback signals may indicate behavioral change is needed; in that case control needs to be exercised in order to successfully *adapt* behavior.

In the context of complex social environments, the ability to control the expression of emotional tendencies in the service of goal achievement represents a particularly important skill. For instance, in a social context, individuals need to be able to inhibit appetitive or angry behavior (Blair & Cipolotti, 2000). In general, studies show gradual increases in the capacity for self-regulation through adolescence, with gains continuing into young adulthood (Steinberg et al. 2008). These developmental improvements of self-regulation are shown to be related to increases in social competence or adaptive social behavior (Kopp, 1982; Nelson et al., 2005; Steinberg, 2009). For instance, the increase in prosocial behavior across adolescence is related to an increased capacity to suppress selfish impulses and forgo short term benefits, in order to acquire the long term benefits of cooperative behavior (Steinberg, 2009). Although, developmental change in self-regulation is often attributed to an increase in the strength of regulatory systems involved in the adaptation of

behavior (e.g. for the suppression of selfish impulses), it may also be attributed to the maturation of *monitoring* processes (e.g. tracking the dynamic changes in the social environment). Finally, similar to perspective-taking, the capacity to self-regulate will not necessarily lead to increased prosocial behavior. The need to self-regulate is dependent on the internal or external goals, and these might be set to achieve anti-social ends.

In sum, the developing capacities for perspective-taking and self-regulation are important factors in the developmental changes in prosocial behavior across adolescence. Although these skills are in and of themselves neutral, an increase in either of these skills is generally positively related to prosocial behavior during normative adolescent development.

It is the hypothesis that structural changes in the developing brain are associated with functional changes in brain networks underlying perspective-taking and self-regulation, and that these changes make an important contribution to the development of adolescent social behavior (Nelson et al., 2005; Steinberg, 2005; Ernst et al., 2008). The next section will summarize recent findings on adolescent structural brain development, and subsequently present recent neurodevelopmental models that will provide a framework for understanding the link between brain changes and changes in behavior.

1.3 Adolescent Brain Development

Early studies of post-mortem brain tissue revealed that the prefrontal cortex of the human brain still shows great changes in synaptic development well into the adolescent period (Huttenlocher, 1979). Additionally, Huttenlocher's work has shown that different areas in the brain show different developmental trajectories in synaptic density (Huttenlocher & Dabholkar, 1997). For instance, the synaptic density of the auditory cortex stabilizes at around age twelve, whereas the prefrontal cortex showed development until at least mid-adolescence.

Grey matter as measured with MRI is proposed to represent the cell bodies, synapses, unmyelinated axons and neuropil. The developmental pattern of grey matter is thought to reflect, at least in part, the processes of synaptogenesis followed by synaptic elimination, or pruning (Huttenlocher, 1979). Several studies have reported a non-linear 'inverted -U' shaped pattern of grey matter development (Giedd et al., 1999; Shaw et al., 2008; Gogtay & Thompson, 2010). The general pattern of grey matter development shows an increase across the cortex prior to puberty, followed by a post-puberty decline. The rise and decline in grey matter follows non-linear patterns and varies depending on the region. The first to mature are the sensorimotor regions, followed by other parts of the cortex in a posterior to anterior direction, with the prefrontal cortex being

one of the last areas to develop (Gogtay & Thompson, 2010). Furthermore, the developmental trajectories of GM vary also within the prefrontal cortex (Gogtay & Thompson, 2010), which could account for differences in rate of development of different control functions associated with these areas (Crone et al., 2006). Correlational studies have shown that differences in prefrontal grey matter volume are associated with individual differences in (anti-)social behavior (Sterzer et al., 2007), suggesting that local quantities of grey matter density may be related to the regulation of social behavior.

In contrast to grey matter, white matter development follows a more linear trajectory, increasing in volume and density during the first two decades of life (Paus et al., 2001). Increases in white matter volume have often been associated with the myelination of axons, but recently it has also been suggested that this could be an effect of increases in axon caliber (Paus, 2010). Both myelination and increases in axon caliber are thought to be associated with increases in processing speed. Studies which focused on structural connectivity using diffusion tensor imaging (DTI) demonstrated that there are still large changes in the fiber tracts that link different brain regions, particularly a rewiring of subcortical-cortical and a strengthening of cortico-cortical connectivity (Supekar et al., 2009; Schmithorst & Yuan, 2010). These connectivity measures have been related to individual differences in adolescent risk-taking (Berns et al., 2009), impulsivity (Olson et al., 2008) and resistance to peer influence (Paus et al., 2008). Thus, there is robust evidence that besides changes in cortical grey matter there are also relations between white matter/structural connectivity and individual differences in traits or behaviors (Cohen et al., 2009).

This brief review showed that there are still substantial changes in brain structures during adolescence[4]. Importantly, some studies showed a relationship between structural differences and individual differences in behavior. This raises the question how changes in brain structure relate to changes in behavior. Although the relation between changes in brain structure and brain function is currently not well understood, most of the current developmental models hypothesize that these structural changes contribute to the development of adolescent behavior via changes in brain function (Nelson et al., 2005; Steinberg, 2005; Ernst et al., 2008; Johnson et al., 2011).

[4] It is important to note that these developmental changes in brain structure are considered to be the result of an interaction between genetic programs and experience. When viewing the images of structural changes in the brain it is tempting to interpret them as the result of a genetically predetermined building plan but in many cases this is incorrect. Take for example synaptic pruning; during this process the synaptic connections that are used are kept and those that are not used are pruned, thus, the result is strongly determined by environmental input and behavior.

Frameworks for understanding development of brain and behavior

Most of the earlier models that were inspired by the novel findings of developmental MRI research hypothesized that the regional differences in structural brain development result in separable developmental trajectories of the specialized functions related to these brain areas. According to this framework brain areas are considered mature when they show an adult pattern of functional activity. Furthermore, because structural development shows linear as well as non non-linear patterns, these models predict linear and non-linear developmental changes in brain function and cognitive skills[5].

In support of these models, the earliest studies on the development of brain function have shown to broadly parallel the findings of structural brain development. Most of these studies that have used functional MRI (fMRI), a technique that makes it possible to examine brain functioning in vivo while participants are performing certain tasks. Consistent with the predictions of the earlier models, neuroimaging research on developmental populations has shown that children and adolescents often use the same network of areas as adults, but that the levels and extent of activity may differ between age groups (Casey et al., 2005). Furthermore, these studies indicated that those brain areas that showed the latest structural development also showed prolonged patterns of functional development.

However, over the past decade several results appeared that seem difficult to reconcile with these models. First of all there are brain areas that show an adult pattern of activity at a very early age in one task but not in others (Bunge & Crone, 2009). Related to these findings are areas that show adult patterns of activation long before they would be considered anatomically mature (for review see Johnson, 2011). Additionally, a strand of research using novel network modeling techniques have investigated the developmental trajectories of several brain networks. These studies on the development of large-scale functional brain networks have shown that in general short-range connections become weaker (segregation) and long-range connections become stronger (integration) with age (Fair et al., 2008; Kelly et al., 2008; Supekar et al., 2009). These findings emphasize that current neurodevelopmental models should take also into account the importance of interregional connectivity, next to the maturation of intraregional connections.

One model that addresses these issues, the model of interactive specialization, assumes that functional brain development involves a process of organizing patterns of interregional interactions (Johnson, 2005; 2011).

[5] Note that most of these theories agree that developmental changes in brain structure are considered to be the result of an interaction between genetic programs and experience, and therefore that functional changes are also a function of both intrinsic and external factors.

According to this view, the response properties of a specific region are partly determined by its patterns of connectivity to other regions and, in turn, by the patterns of activity of these other regions. During development, activity-dependent interactions between regions results in functional specialization of areas and networks, which may be reflected in both regional changes in activation patterns and changes in connectivity between regions. Thus, because the function of a brain region is co-determined by its place in a network its pattern of activity is also dependent on: (1) the strength of the connectivity with other areas, and (2) the level of activity in these other areas. As a result, it is possible that in certain situations an area may show similar levels of activations for children and adults, but not in others[6].

Additionally, specialization is thought to result in cortical regions or networks becoming more specialized in their response properties; they will therefore respond less to the non-preferred stimulus or task contexts with increasing age. This specialization process may be reflected in changes from distributed to focal activation of certain brain areas with age (Durston et al., 2006), or in the number of areas that are activated within a certain network (Scherf et al., 2006; Johnson, 2010).

From this developmental framework follows the prediction that developmental changes in social interactions are related to (1) regional changes in activation patterns and (2) changes in connectivity between regions or networks that underpin perspective-taking and self-regulation. Before turning to a more detailed description of the neural underpinnings of perspective-taking and self-regulation, the question how to experimentally study the development of social interactions is addressed.

[6] This view even further complicates inferring cognitive function from brain activity. As Russel Poldrack (2006) argued 'reverse inference' from brain activation to cognitive function (2006) is not deductively valid, but rather reflects the logical fallacy of affirming the consequent. Furthermore, Poldrack proposes a Bayesian approach for estimating the likelihood a specific function is associated with a specific brain area. However, according the IS framework it is possible that a certain areas functions differently at different stages of development due to its changing connections to other areas. That makes 'regressive reverse inference' even more dangerous because even if there is substantial evidence for cognitive function A being related to activation in area B in adult studies this does not tell us how probable it is that cognitive function A is also related to function B in children. Thus we should be extra cautious with 'regressive reverse inference', however, as Poldrack suggested we should not completely refrain from it. And in an emerging field, such as developmental neuroimaging, it is probably a necessary evil. Nevertheless, it emphasizes the need for strong theoretical predictions when interpreting neurodevelopmental data, because these are currently the best protections against misinformed inferences.

1.4 Studying Social Interactions

Previous research in developmental psychology on perspective-taking and prosocial behavior is mainly based on self- or other-reports. (e.g. parents, peers or teachers) Currently, there are a few studies which have shown that there are subtle developmental changes on experimental measures of perspective-taking during adolescence (e.g. Choudhury et al., 2006; Duhmontheil, et al.2009), but these studies did not examine perspective-taking in a social context. Furthermore, the correlations between perspective-taking skills and prosocial behavior are stronger for self or other-report indices than for responses to hypothetical social scenarios (Eisenberg & Schell, 1986). Importantly, this suggests that the relation between perspective-taking and social behavior is best studied using real social interactions rather than hypothetical social scenarios (Gummerum, Hanoch & Keller, 2008). Similarly, most of the research on the development of self-regulation is based on questionnaires, and although there are some experimental measures, these are not related to social interactions (Steinberg, 2009).

The challenge is therefore to find experimental paradigms which allow for the study of the development of perspective-taking and self-regulation in the context of social interactions. Additionally, for the purposes of this thesis, these paradigms also needed to be suitable for both developmental populations and the constraints of MRI research.

To investigate the psychological and neural correlates of prosocial behavior in social interactions, the experiments in this thesis are based on Game Theoretical paradigms (economic games) derived from experimental economics (Neumann & Morgenstern, 1947) and social psychology (Camerer, 2003; Sanfey, 2007). In these experiments participants interact with other people in simple bargaining or exchange games with real monetary consequences. These games often simulate single interactions between two anonymous individuals, focusing on the motivations of prosocial behavior. However, the games can also consist of multiple interactions over time in order to study the regulation of social behavior in a dynamic environment (Kishida et al., 2010).

The advantage of economic games is that their structural simplicity yields precise characterizations of complex social behavior, which makes the paradigms also suitable for neuroimaging experiments. A second strength of games is that behavior can be operationalized in the same way across age groups (Gummerum et al., 2008). Finally, the ecological validity of these games has been well assessed in prior work (for a review, see Camerer, 2003). For example, prosocial behavior in these games is predicted by participants' actual prosocial behavior in the past (Glaeser et al., 2000) and by their estimation of their expected prosocial attitude in real-life situations (van Lange et al., 1997).

Next, two economic games, the Trust and Ultimatum Game, will be presented in more detail. These two games are used often in the neuroimaging literature, and zoom in to several important aspects of prosocial interactions: trust, reciprocity and fairness.

Trust, Reciprocity & Fairness in Economic Games

In the Trust Game (Berg et al., 1995) two players can share a certain amount of money. The first player can choose to divide the money equally between herself and the second player, or to give it all to the second player with the advantage that the stake increases in value (see Figure 1.1). The second player has the choice to reciprocate and share the increased amount of money with the first player, or to defect and exploit the given trust by keeping the money for herself. As a result, the Trust Game models both the decision to trust and to reciprocate trust. The Trust Game can be played a single interaction or as an iterated multiple-round game. Game Theory predicts that the second player in a single interaction Trust Game would never reciprocate the trust given by the first player because taking all the money will have no negative future consequences. Taken this into account, the first player will therefore also never trust the second player. In contrast with these predictions experimental data show that most people trust the second player, and also that the second player's trust is generally reciprocated (Camerer, 2003).

In iterative multiple round Trust Games, when the same participants interact over a number of rounds, the theoretical predictions and actual behavioral strategies change (Axelrod, 1984). Studies with multiple round games have shown that participants often play a tit-for-tat like strategy (Wedekind & Milinski, 1996; Nowak & Sigmund, 1992). That is, if the second player shared in the previous round the first player will trust in the following round (positive reciprocity), and if the second player did not share in the previous round the first player will react by not trusting in the next round (negative reciprocity). Because of their dynamic nature, these types of games are a useful tool for investigating the processes involved in the monitoring and regulation of social interactions (King-Casas et al., 2008; Kishida et al., 2010).

The second economic game of interest, the Ultimatum Game (Guth et al., 1982), is a bargaining game in which the first player (proposer) is given a sum of money to share with the second player (responder). If the responder accepts the amount offered by the proposer, the money is split between the two as proposed. However, if the responder considers the proposed split unfair and rejects the offer, neither player receives any money (See Figure 1.1). Game Theory predicts that responders will always accept offers that are larger than zero, because rejecting would leave them with less. However, on average,

responders already start rejecting offers less than 40% of the stake, suggesting that their decisions are not only driven by material interests but are also based on self–other comparisons, or "fairness considerations" (Straub & Murnighan, 1995).

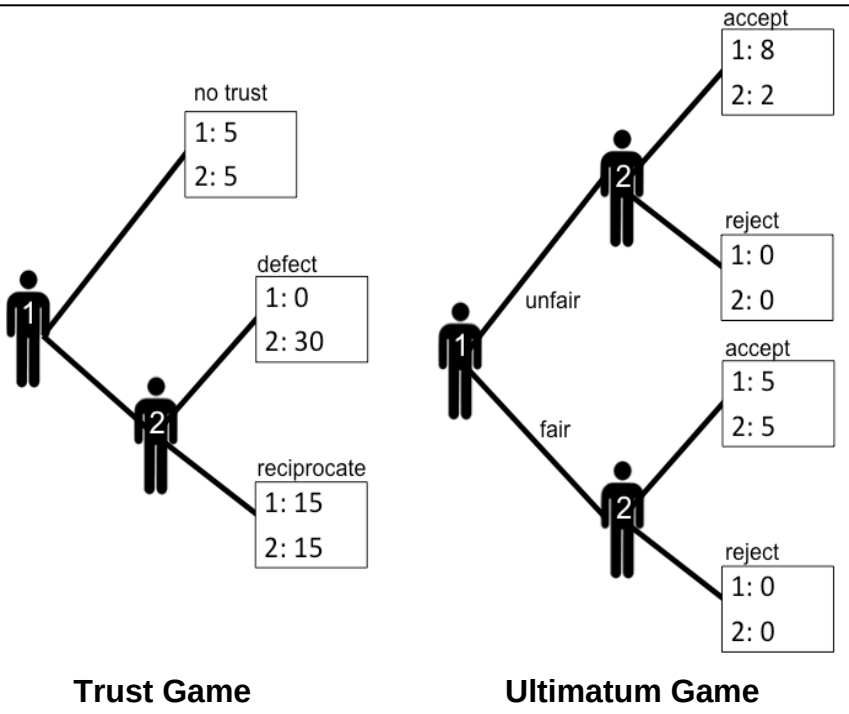


Figure 1.1 An example of the Trust Game and the Ultimatum Game. The graph of the Ultimatum Game represents two possible offers that a first player can make: an unfair (8 for self 2 for the other) and a fair (equal split, 5 for each) offer. If the second player rejects both players end up with nothing, regardless of the offer.

To conclude, these two simple games have been successfully applied in many studies to investigate different aspects of prosocial behavior in social interactions. The one-shot single interaction Trust and Ultimatum games have proven to be useful to study the role of perspective-taking in social decision-making (Pillutla, et al., 2003; Malhtora et al., 2004; Sutter, 2007; Falk et al., 2008), whereas the multiple-round versions of these games are useful for studying the monitoring and adaptation of social behavior in dynamic social environments (Delgado et al., 2005; King-Casas et al., 2005; Krueger et al., 2007; Behrens et al., 2009). Both types of games are therefore able to capture the processes of interest related to the development of social behavior. The next

section reviews what recent neuroimaging studies have revealed about neural mechanism of social decision-making in adults.

1.5 Social Decision-Making and the Brain

To start, one important insight from neuroimaging studies with economic games is that social decision-making is the product of multiple interacting systems (Sanfey, 2007; Behrens et al., 2009; Frank et al., 2009). This section will focus on those networks of which the associated functions are related to the cognitive processes identified earlier as underlying developmental changes in social behavior, namely perspective-taking and self-regulation. First, studies of social interactions emphasizing the importance of a specific 'social brain' network (Amodio & Frith; 2006; Hampton et al., 2008) will be addressed. These are followed by a review of social interaction studies that have emphasized the importance of brain regions with a general role in monitoring (Delgado et al., 2005; King-Casas et al., 2005) and regulating behavior (van 't Wout et al., 2005; Knoch et al., 2008). The function of the different brain networks will be discussed in the context of economic games, followed by a review of recent evidence of developmental changes within these networks.

First, the 'social brain' network (Frith & Frith, 2003; Van Overwalle, 2009), thought to be involved in thinking about other people's believes and intentions, consists of the anterior medial prefrontal cortex (aMPFC), temporal poles (TP), posterior superior temporal sulcus (pSTS) and the temporal parietal junction (TPJ, see Figure 1.2). Prior neuroimaging studies have shown that specifically the aMPFC and the TPJ are involved in processes related to perspective-taking. For example, neuroimaging studies have demonstrated that aMPFC and TPJ are active during theory-of-mind tasks, such as tasks that require participants to infer mental states of characters in stories (Fletcher et al., 1995) and cartoons (Gallagher et al., 2002) or while watching animations (Castelli et al., 2000). In addition, prior studies have suggested that in context the context of social interactions the aMPFC is involved in evaluating the mental content of others in relation to the self (Amodio & Frith, 2006), whereas the TPJ is thought to be important for redirecting or focusing attention on the other (Saxe et al., 2004; Mitchell, 2008; Hampton et al., 2008; Blakemore, 2008; Van Overwalle, 2009). For instance, aMPFC activity has been reported when participants trust another individual, with the expectation of increasing their own pay-off (McCabe et al., 2001). On the other hand, the TPJ activity was increased when participants considered the intentions of other player in a competitive game (Halko et al., 2009). These results suggest that in social interactions the aMPFC is important

for the evaluation of own outcomes, whereas TPJ activation may indicate a focus on the outcomes of others.

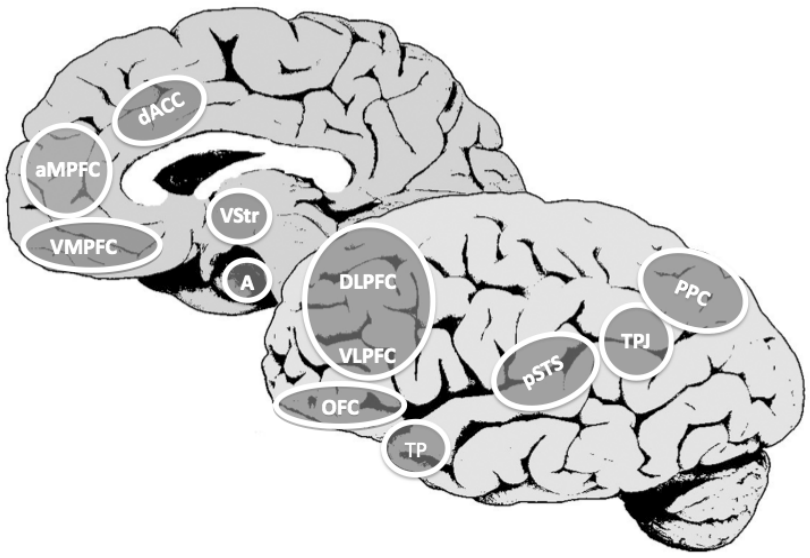


Figure 1.2 Schematic representation of the networks of brain areas involved in social decision-making: aMPFC = anterior Medial Prefrontal Cortex, TPJ = Temporal Parietal Junction, pSTS = posterior Superior Temporal Sulcus, TP = Temporal Poles, Vstr = Ventral Striatum, A = Amygdala, VMPFC = Ventro Medial Prefrontal Cortex OFC = Orbito frontal Cortex, dACC = dorsal Anterior Cingulate, DLPFC = Dorsolateral Prefrontal Cortrex, VLPFC = Ventrolateral Prefrontal Cortrex, PPC = Posterior Parietal Cortex.

Second, this 'social brain' network is found to work together with areas that are involved in the regulation of social behavior. This regulatory network includes the lateral prefrontal cortex (VLPFC & DLPFC), dorsal anterior cingulate cortex (dACC) and the posterior parietal cortex (PPC) (Botvinick et al. 2001, Miller & Cohen, 2001, Pochon et al. 2008). In the context of social interactions it is thought that the dACC is involved in signaling a conflict of interest. For instance, the dACC is more active in case of an unfair offer in the ultimatum game, when there is a conflict between social norms (it is unfair) and personal interest (it is an amount of money for me) (Sanfey et al., 2003). Other studies have shown that the DLPFC is also more active when a norm is violated, such as an unfair offer in the Ultimatum Game (Sanfey et al., 2003; van't Wout et al., 2005; Knoch et al., 2006; Tabibnia et al., 2008). It has been suggested that in those cases when there is conflict between different motivational drives the higher level control areas, such as the DLPFC, have a role in regulating social behavior (Sanfey, 2007; Frith & Singer, 2008). In this example, the DLPFC is

thought to have a role in controlling the selfish impulse to accept a small amount of money, in order to (costly) punish the other player for violating a social norm (Knoch et al., 2008).

Additionally, there is a network that includes the striatum, ventral medial prefrontal cortex (VMPFC) and insula, which is involved in monitoring the behavior of self and others in multi-round games (Delgado et al., 2005; King-Casas et al., 2005; Krueger et al., 2007; Behrens et al., 2009). The neuroimaging studies of social interactions showed that the insula is engaged when social norms are violated. For example, the insula shows increased activity during unreciprocated trust (Montague & Lohrenz, 2007; Rilling et al., 2008) and unfair proposals in the Ultimatum Game (Sanfey et al., 2003; Tabibnia et al., 2008). In contrast, striatum and VMPFC activity correlate positively with cooperation choices in the Trust Game (Rilling et al., 2004; Krueger et al., 2008). Thus, these areas seem to be involved in signaling *and* learning the pleasant and unpleasant aspects of social interactions, which may explain how lower level affective processes contribute to the regulation of social behavior (Sanfey, 2007).

Social Decision-Making and the Developing Brain

Finally, although the development of these networks has not been studied specifically in the context of social interactions, age related changes in brain function have been observed in each of these networks separately.

First, a series of studies have investigated functional changes in the social brain network during adolescence in passive social paradigms that involved thinking about self and others. In general, these studies showed an age related shift in activation from the aMPFC to the TPJ (for review see Blakemore, 2008). It can be hypothesized that this age related shift in pattern of activity reflects that early adolescents still rely more on self-reflective processes performed by the aMPFC, whereas later in adolescence they are more engaged in other-focused processes performed by the TPJ. This shift in processing of social stimuli is consistent with descriptive theories that suggest an important relation between a shift in perspective-taking and the development of prosocial behavior during adolescence (Eisenberg et al., 1995, 2001).

Second, developmental studies of performance monitoring have shown that networks involved in the regulation of behavior, such as the ACC and DLPFC develop until late adolescence. First, age related increases in the error-related negativity (ERN), a scalp potential thought to reflect dACC activity, are consistently reported across studies (Davies et al., 2004; Ladouceur et al., 2004). These changes in the ERN are suggested to reflect an age related increase in the ability to monitor feedback signals and regulate subsequent

behavior. Consistent with these results, recent neuroimaging studies have shown age related changes in dACC, DLPFC and PPC in performance monitoring until late adolescence (Crone et al., 2008; van Duijvenvoorde et al., 2008). Given the role of the these areas in adult social decision-making, it can be hypothesized that the reported developmental changes in brain activation contribute to the ability to monitor and regulate social behavior in relation to internal and external goals (e.g. personal and social norms).

In sum, adult neuroimaging studies of social interactions have shown that there are multiple networks of brain areas that are related to the capacities of perspective-taking and self-regulation in social decision-making. Furthermore, there is evidence that developmental changes take place in these networks until late adolescence. From these results follows the prediction that the developmental changes in perspective-taking will be related to the function of the 'social brain' network, whereas developmental changes in self-regulation are expected to be found in networks involved in monitoring and regulating social behavior. Furthermore, the prediction from these studies, and neuro-developmental models, is that developmental outcomes are also the result of the interplay between these different networks. The experiments described in this thesis aimed at investigating the developmental changes in functional activity, and connectivity *within* these networks, to further our understanding of the mechanisms underlying social development across adolescence.

1.6 Outline of current thesis and publications

In the first empirical chapter (**Chapter 2**) a child friendly version of the Trust Game is developed. The Developmental Trust Game (DTG) is a Trust Game with outcome manipulations that allowed testing the sensitivity for the perspectives of others. In the following chapter (**Chapter 3**) the DTG paradigm was used to explore the neural correlates of reciprocating trust in relation to individual differences in social value orientation and perspective-taking manipulations in an adult population. The two subsequent chapters describe developmental changes in neural correlates of perspective-taking in reciprocal behavior (**Chapter 4)** and fairness judgments (**Chapter 5**). In **Chapter 6** a child friendly behavioral paradigm to study developmental changes in multiple social interactions is introduced; the Simultaneous Trust Game (STG). The two subsequent studies have investigated the neurodevelopmental changes of feedback processing while performing a probabilistic learning task. The first study (**Chapter 7**) investigated the developmental changes in feedback processing in the context of learned rules, focusing on the dACC, PCC and DLPFC network, whereas the second study (**Chapter 8**) investigated the

neurodevelopmental changes of feedback during the learning phase in the same task. The second study focused on the developmental changes in connectivity strength within the striatum-medial prefrontal network. Although the latter two studies of feedback processing are not conducted in the context of social decision-making paradigms, they provide important building blocks for interpreting the developmental changes in the processes underlying self-regulation in social behavior. In the final chapter (**Chapter 9**) the results of the empirical studies will be summarized and discussed

The following papers have resulted from this thesis:

van den Bos, W., & Crone, E.A. (to appear in 2011) The Neuroscience of Social Decision-Making: A Developmental Perspective. In *'Neural Basis of Motivational and Cognitive Control'* (R. Mars, J. Sallet, M. Rushworth, & N. Yeung, eds.). MIT press. **(Chapters 1 & 9)**

van den Bos, W., Westenberg, P.M., van Dijk, E. & Crone, E.A. (2010) Development of Trust and Reciprocity in Adolescence. *Cognitive Development*. 25 (1), 90-102. **(Chapter 2)**

van den Bos, W., van Dijk, E., Westenberg, P.M., Rombouts, S.A.R.B. & Crone, E.A. (2009), What motivates repayment? Neural correlates of reciprocity in the Trust Game. *Social Cognitive and Affective Neuroscience*. 4(3), 294-304. **(Chapter 3)**

van den Bos, W., Van Dijk, E., Westenberg, P.M., Rombouts, S.A.R.B. & Crone, E.A. (2010) Changing Brains, Changing Perspectives: The Neurocognitive Development of Reciprocity. *Psychological Science*. **(Chapter 4)**

Güroğlu, B., **van den Bos, W**., Rombouts, S.A.R.B., & Crone, E.A. (submitted) Dissociable brain networks involved in development of fairness considerations. **(Chapter 5)**

van den Bos, W., van Dijk E., & Crone, E.A. (submitted) Who do you trust? Age comparisons of learning who to trust or distrust in repeated social interactions. **(Chapter 6)**

van den Bos, W., van der Bulk, B. G., Güroğlu, B., Rombouts, S.A.R.B. & Crone, E.A. (2009) Better than expected or as bad as you thought? The neurocognitive development of probabilistic feedback processing. *Frontiers in Neuroscience*, 3, 52 **(Chapter 7)**

van den Bos, W., Cohen, M.X., Kanht, T., & Crone, E.A. (submitted) Striatum-medial prefrontal cortex connectivity predicts developmental changes in reinforcement learning. (**Chapter 8)**

2. Development of trust and reciprocity in adolescence

We investigate the development of two types of prosocial behavior, trust and reciprocity, as defined using a game-theoretical task that allows investigation of real-time social interaction, among 4 age groups from 9 to 25 years. By manipulating the possible outcome alternatives, we could distinguish among important determinants of trust and reciprocity that are related to the risk and benefit of trusting. The results demonstrate age related changes in sensitivity to outcome for others from late childhood until late adolescence, with different developmental trajectories for trust and reciprocity and differential sensitivity to risk and benefit for self and others.

2.1 Introduction

Adolescence is a developmental period characterized not only by physical and hormonal changes but also by substantial changes in social behavior (Steinberg, 2005). Most notable is change in the nature of social interactions, from competitive to more prosocial behavior (Eisenberg, Carlo, Murphy, & van Court, 1995; Eisenberg, Miller, Shell, McNalley & Shea, 1991; O'Brien & Bierman, 1988; Schaffer, 1996; Van Lange, Otten, de Bruin & Joireman, 1997). Developmental theorists suggest that a prosocial attitude develops during adolescence as a part, or as a consequence of, the development of increased capability for social perspective-taking (Eisenberg et al., 1991, 1995; Kohlberg, 1981; Selman, 1980).

With development, adolescents learn to better understand the perspective of the other and to coordinate between the different perspectives of self, others and society (Martin, Sokol & Elfers, 2008).Perspective-taking is a complex, multi-factor construct; yet there is evidence for at least a weak correlation between perspective-taking and prosocial behavior in adolescence (Underwood & Moore, 1982). Notably, these correlations are stronger for self-report indices than for responses to hypothetical scenarios of prosocial behavior (Eisenberg & Schell, 1986), suggesting that prosocial behavior is best studied using real-life rather than hypothetical social scenarios. Here, we study the development of

prosocial behavior using a two-person interaction game, and we define perspective-taking as the ability to consider outcomes for self in relation to outcomes of others.

Game-theoretical studies can provide an authentic social interaction context in which a 'theory of mind in action' can be investigated experimentally (Gummerum, Hanoch & Keller, 2008). In contrast to studies involving hypothetical scenarios, decisions in games have real consequences. Players allocate real money between themselves and the other player and are paid according to their decisions.Consequently, behavior in games may be more similar to that in real-life contexts. Another strength of using games as a measure of prosocial behavior is that behavior can be operationalized in the same way across age groups (Gummerum et al., 2008). One such game, the Trust Game (Berg, Dickhaut, & McCabe, 1995), is of particular interest for understanding the changes in social cognition that occur during adolescence because it allows us to separately examine two important types of prosocial behavior, trust and reciprocity.

Trust and reciprocity can be considered key elements of prosocial behavior. Proscocial behavior is often characterized by exchanges of favors between non-related individuals (Camerer, 2003). Often these exchanges of favors are separated in time, such that a favor will only be returned on a future occasion. Trust in positive reciprocity at future times is therefore essential to initiate a cooperative interaction. Additionally, reciprocity is necessary to maintain social relationships; if favors are not returned relationships may be short-lived (Lahno, 1995).

In the Trust Game, two anonymous players are involved in dividing an amount of money. The first player, the trustor, has the possibility of dividing a certain amount of money between self and other. However, the trustor can also decide to give all the money to the other who then is able to divide the money; in that case the total amount that is divided between the two players increases. If the second player gets the chance to decide how the money is divided, he or she is confronted with two options—to equally share the money (reciprocate) or to keep most of the money and to give only a small amount to the first player (exploit)[7]. As a consequence, the first player has the possibility of gaining more money if he or she decides to give the money to the second player. However, in doing so the first player also takes the risk that the second player will not reciprocate. Typical findings in the Trust Game are that adults often choose to trust and reciprocate, even when doing so is potentially costly (Berg et al., 1995;

[7] Following Malhotra (2004), we use the terms 'reciprocate' and 'exploit' to describe the two options of player 2. Other common terminology is 'honoring trust' versus 'abuse of trust' (e.g., Buskens, 2003). Note that these labels were not used to explain the paradigm to the participants.

Bolle, 1995; Dufwenberg & Gneezy, 2000; Ortmann, Fitzgerald, & Boeing, 2000; McCabe, Houser, Ryan, Smith, & Trouard, 2001).

In this study, we examine the development of trust and reciprocity in the context of social interaction with anonymous others in the Trust Game. The study is different from studies in which the social interaction examined is with friends, peers or parents (Bernath & Feshbach, 1995; Brett & Willard, 2002; Laursen & Hartup, 2002; Rotenberg et al., 2005; Youniss, 1980). The anonymous method allowed us to examine amore generalized form of trust and reciprocity, underlying all forms social interactions (Rotenberg et al., 2005). The ecological validity of these games has been well assessed in prior work (for a review, see Camerer, 2003). For example, trust behavior in the Trust Game has been shown to be predicted by participants' actual trust behavior in the past (Glaeser et al., 2000) and by their estimation of reliability in real-life situations (Rotenberg et al., 2005).

A prior developmental study using the Trust Game has demonstrated an increase in trust and reciprocity with increasing age among participants of 6 age groups (8, 12, 16, 22, 32, and 68 years; Sutter & Kocher, 2007). With age, participants offered more money and also returned more money; this behavior stabilized between 16 and 22 years of age.

Both trust and reciprocity as defined here are hypothesized to require social perspective-taking abilities, in order to recognize the intentions of the trustor and predict whether the trusted person is likely to reciprocate (Pillutla, Malhotra & Murnighan, 2003; Malhotra, 2004). Based on the theoretical framework that presupposes a relation between development of prosocial behavior and social perspective-taking (Martin et al., 2008), our goals were to investigate the processes related to perspective-taking that may account for changes in trust and reciprocity and to identify the developmental trajectories.

To address these questions, we developed a developmentally appropriate version of the Trust Game (Berg et al., 1995), the Developmental Trust Game (DTG). The DTG is presented in a computerized format and is appropriate for younger participants because the monetary amounts players must divide between themselves are represented with coins instead of numbers and the amounts are relatively small (1–20). The task thus poses a similar level of cognitive difficulty for the youngest children and for late adolescents (for other examples, see Crone & van den Molen, 2004). As in prior studies with adults (Malhotra, 2004), we presented participants with a fixed two-choice paradigm, in which player 1 (the trustor) has the possibility to either trust or not trust the other player. Player 2 (the trustee) also has two choices, to reciprocate and divide money about equally, or to exploit and keep most of the money (see Fig. 2.1).

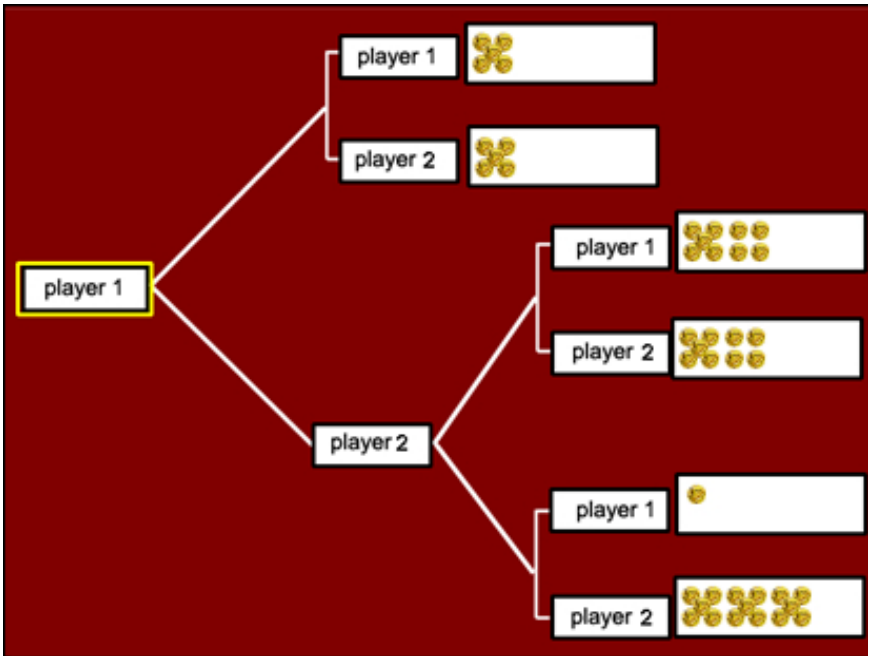


Figure 2.1: The sequence of visual displays that represent the different stages of the Children's Trust Game (DTG). Each round starts with the identification of each player and the designation of player 1 and player 2. Then the game starts and each player is presented with the complete decision tree. At this point player 1 can decide not to trust and the game ends. If player 1 decides to trust, the box with the name of player 2 is highlighted. Subsequently player 2 ends the game by either exploiting or reciprocating. The stakes are represented by a number of 10-cent coins in boxes next to the names of the players. Note that the labels on this figure are for illustrative purposes only and were not visible to the participants.

To examine the role of perspective-taking, defined as the ability to consider the intentions of and consequences for others, we added experimental manipulations to the original Trust Game that may reveal whether participants are taking the intentions of others and consequences for others into account (Pillutla et al., 2003; Malhotra, 2004). We manipulated two factors that may affect trust and reciprocity decisions: the risk of making a decision to trust (risk) and the extent to which a decision to trust is beneficial to the trustee (benefit). Therefore, this design has the potential to reveal more specific developmental changes relative to reports on the average levels of trust and reciprocity among different age groups.

Risk, benefit and perspective-taking: developmental paths in trust and reciprocity.

Trusting always involves a certain amount of risk. When a favor is provided, there is always a chance that it will not be reciprocated. Following Malhotra (2004), we therefore manipulated risk for the trustor by varying the outcome that player 1 can obtain if player 1 decides not to trust player 2 (see Fig. 2.2).

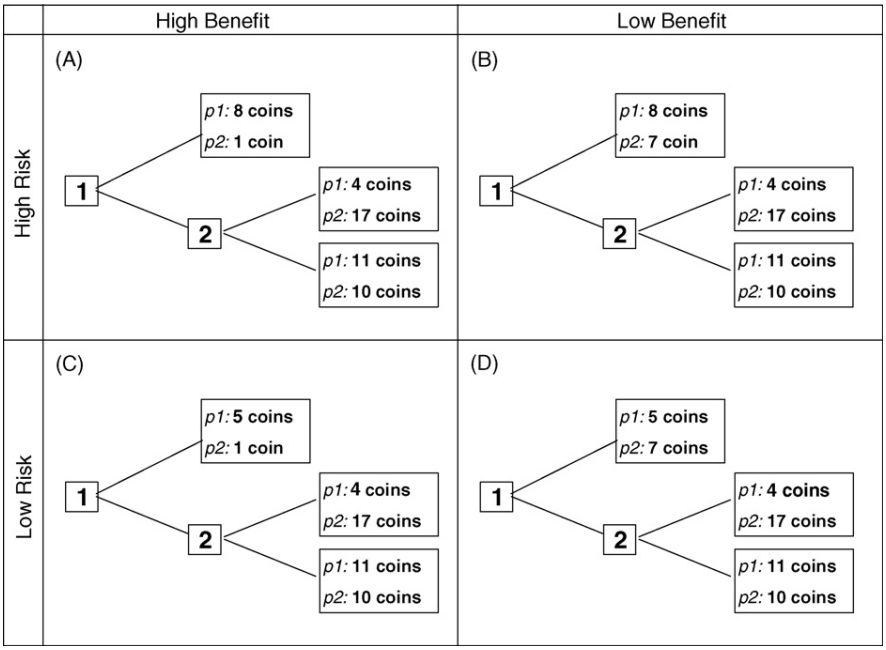


Figure 2.2: Visual representation of the 4 experimental conditions: (A) high risk and high benefit, (B) high risk and low benefit, (C) low risk and high benefit and (D) low risk and low benefit.

In the high-risk conditions, player 1 ensures a high outcome by deciding not to trust player 2. A decision to trust player 2 means that player 1 takes a high risk by forfeiting assurance of this high outcome. In the low risk conditions, player 1 stands to gain only a relatively low outcome. A decision to trust player 2 means that player 1 takes only a low risk by forfeiting assurance of a relatively low outcome (Fig. 2.2).

Consistent with Malhotra (2004), who used a similar manipulation to vary risk, we predicted that player 1's trust decisions would be affected by our risk manipulation. Participants should less often opt to trust player 2 when facing a high-risk decision than when facing a low-risk decision. Because the risk manipulation only affects own outcome for the trustor, and therefore does not

require extensive perspective-taking skills, we expected to see a similar effect of increased risk on player 1's decision to trust at all ages.

With regard to player 2's decisions to reciprocate, we did expect age effects. Increased risk for the trustor may increase the amount of reciprocity by the trustee. In that case the trustee will reciprocate the risk taken by the trustor. Note, however, that it requires the trustee to take the perspective of the other in order to recognize the risk taken by the trustor. Because perspective-taking is thought to develop in adolescence, we expected that that the increase of reciprocity with risk would be larger for adults than for younger participants.

In addition to the risk for the trustor, we also considered the extent to which a decision to trust would benefit the trustee (Malhotra, 2004). Being trusted always involves a certain benefit, which one might or might not reciprocate. Following Malhotra (2004),we therefore also manipulated the benefit for the trustee (player 2) by varying the outcome that player 2 obtains if player 1 decides not to trust player 2. In the low-benefit conditions, player 2 is already assured a high outcome if player 1 decides not to trust player 2. A decision to trust player 2, is therefore only of limited benefit to player 2. In the high-benefit conditions, player 2 receives only a relatively low outcome if player 1 decides not to trust player 2. A decision to trust player 2 is therefore highly beneficial to player 2 (Fig. 2.2).

It is important to distinguish between decisions to trust (player 1 decisions) and decisions to reciprocate (player 2 decisions). With regard to decisions to reciprocate, it seems likely that trustees are more likely to reciprocate when the benefit for being trusted is higher. In other words, we anticipated that participants would value the fact that the trustor takes their benefit into account by subsequently reciprocating. Note, however, that for the trustee to recognize that the trustor took their benefit into account requires perspective-taking. Furthermore, we predict that trustors are more likely to trust when the benefit for the trustee is higher, anticipating the previously proposed increased generosity. Note again that this effect requires the trustor to take the perspective of the trustee; it requires making an inference of the effect of benefit on the state of mind, and subsequent behavior, of the trustee. Thus, in contrast to the risk manipulation, an effect of benefit always requires a certain amount of perspective-taking for both trustor and trustee. Therefore, we expect high benefit to lead to an increase in trust and reciprocity. We expect this benefit effect to be stronger for adults and possibly even absent for the youngest participants.

In addition to the manipulation of benefit and trust we included a control condition to make sure that participants of all ages, especially the youngest, understand the structure of the game. In the control condition it was always best

to trust and to reciprocate, because this would lead to the highest gains for both parties. Therefore, we expect no age differences in trust or reciprocity in the control condition.

We designed the experiment such that participants played multiple games as both trustor and trustee. This design allowed us to examine both trust and reciprocity in the same individual. Importantly, participants were instructed that they were always coupled with a different player.

2.2 Method

2.2.1 Participants

Our sample included 92 participants (49 male) in four age groups: late childhood (M age = 9.43, SD = .59, 12 male, 11 female), early adolescence (M age = 12.35, SD = .56, 17 male, 9 female), middle adolescence (M age = 15.65, SD = .58, 9 male, 14 female) and late adolescence (M age = 22.3, SD = 2.4, 11 male, 9 female). Chi-square analyses indicated that gender distributions did not differ significantly by age. Children and adolescents were recruited from local schools. Adults were university students.

Participants were selected from schools whose populations have common Dutch ethnicity and were mostly Caucasian. Child and adolescent participants were selected with the help of their teachers (children with learning or psychiatric disorders were excluded); informed consent was obtained from a primary caregiver.

2.2.2 Developmental Trust Game

The Developmental Trust Game (DTG, Fig. 2.1) is a version of the Trust Game (Berg et al., 1995; Malhotra, 2004) appropriate for a wide age range. The DTG presents small amounts of money with a number of 10-cent coins in each box of a decision tree.

In each trial, participants were randomly assigned to the role of player 1 (the trustor) or player 2 (the trustee) by a display that was presented for 2500 ms. This screen displayed the first name and picture of both players. After the roles of the participant and the other player were assigned, the trial started. The other player was always matched for age and gender. Participants were told that a different anonymous individual would be paired with them at each trial. However, they actually played against a computer simulation.

Player 1: Trustor. When the participant was assigned the role of player 1 (trustor), the task involved two steps. First, at the beginning of the trial the participant saw the complete decision tree and had to choose between two options: to trust or not to trust. The whole decision tree was represented such

that the player could always see the risk and benefit for each possible choice. If the participant decided not to trust, the coins were divided between the players as represented by the number of coins in each box. If the participant decided to trust, the number of coins in the game was increased and the control of the outcome was in the hands of player 2 (trustee). The choice of the participant (player 1) was presented on the outcome screen by a change in the color of the boxes. The participant then waited for the choice of player 2. The participant was told that the other player made his or her decisions through an internet connection but in reality the choice was made by the computer program after a variable delay of 2–4 s (see Table 2.1 for computerized response pattern). The presentation of this decision was displayed by changing the color of the box representing the choice of the other player. The presentation of the outcome of the trial was displayed for 3 s.

Player 2: Trustee. When the participant was assigned the role of player 2 (trustee), the task also involved two steps. First, the participant awaited the choice of player 1. The participant was told that player 1 would make a decision through an internet connection. In reality, the choice was made by the computer, and the choice was presented within a 3–5 s interval. At this stage, if player 1 chose to trust, the participant was presented with two options: reciprocate or exploit. If player 2 decided to exploit, player 2 would take most of the money and player 1would get fewer coins than in the no-trust option. If player 2 reciprocated the coins were shared equally and both players received more coins, compared to the no-trust option. Risk for the trustor (high versus low) and benefit for the trustee (high versus low) were manipulated, similar to the paradigm used by Malhotra (2004) (see Fig. 2.2). The risk manipulation determined the risk involved in trusting for player 1. If the risk was low, player 1 could potentially lose a small number of coins by trusting player 2 if player 2 chose to exploit the trust (e.g., a loss of 1 coin compared to the no-trust option, see Fig. 2.2 C and D). In contrast, when the risk was high, player 1 could potentially lose a relatively large number of coins by trusting player 2 (e.g., a loss of 4 coins, see Fig. 2.2 A and B). The benefit manipulation determined the benefit for player 2 of being trusted by player 1. In the low-benefit condition, player 2 would get a large number of coins in the no-trust option; therefore the benefit of being trusted was rather small (Fig. 2.2 B and D). The number of coins for player 2 in the no-trust option in the high-benefit condition was small. As a result, there was a large increase of coins (benefit) for player 2 in the case of trust (Fig. 2.2 A and C). The control condition entailed a decision tree in which the option to trust always resulted in a higher pay-off than the no-trust option, regardless of the choice made by player 2.

Table 2.1. Computer simulations of trust and reciprocity for each condition.

	High Risk		Low Risk	
	Trust	Reciprocate	Trust	Reciprocate
High Benefit	47%	73%	60%	67%
Low Benefit	33%	27%	53%	20%

A fixed schedule was used for each of the roles and conditions (Table 2.1), following previous work (Malhotra, 2004). In total, the task consisted of 15 low-benefit–low-risk trials, 15 low-benefit–high risk trials, 15 high-benefit–low-risk trials, 15 high-benefit–high-risk trials, and 10 control trials, for both the trustor role and the trustee role. Consequently, for each participant the task consisted of 140 trials in total. The rounds were presented in random order, and there were breaks after every 20 rounds. The experiment was self-paced and took between 30 and 45 min to complete. At the end of the experiment a screen was presented which displayed the pay-off. The individual pay-off was a variable amount between 3 and 5 Euros. Because previous research with the trust game paradigm has shown that the size of the stakes does not significantly change behavior within different age groups between 8 and 68 years old (Sutter & Kocher, 2007), we were confident to use the same stakes level for all age groups.

2.2.3. Procedure

Child and adolescent participants were individually tested at their school in a quiet room and adult participants were tested in a laboratory, using a standard desktop computer or a laptop. All participants received initial verbal instructions and filled out a questionnaire to assess whether they understood the structure of the game. Subsequently, they played 18 practice rounds to become familiar with the interface. The experimenter personally went over the participant's answers and provided any necessary additional explanation; if necessary an additional set of practice rounds was presented.

Participants were instructed that they were going to play an interactive game with a number of anonymous other players withwhomthey were connected via the internet. It was emphasized that the other participants were unfamiliar to them, coming from other schools or universities participating in the experiment. Only the first name and the first letter of the surname were presented on the screen to identify the other player (e.g. Wouter B.). We used a set of avatars showing silhouettes of real people, instead of real pictures, to prevent their influence on judgments.

Participants were told that at the end of the experiment the computer would randomly select four rounds and the total outcome for the participant in those rounds determined the pay-off. Participants were also reminded that the same rule applied to all the other players they would encounter in the game, to emphasize that their decisions had potential consequences for themselves and others. Participants were paid directly after the experiment. All participants were debriefed at the same time.

Following the DTG, all participants completed the Raven Standard Progressive Matrices (SPM), a non-verbal test of general intellectual ability (Raven et al., 1998). SPM scores were transformed, correcting for age, to IQ estimates. The total duration of the experiment was approximately 65 min.

2.3 Results

2.3.1 Raven SPM

We first examined whether the different age groups differed in general intelligence and the effect of IQ differences on performance. As expected, the number of correct solutions on the Raven SPM task increased with age. Raven scores were z-transformed, using different transformation for different ages, to enable comparisons across age groups. The individuals of all age groups had above average IQs as estimated by transformed Raven SPM scores; 9–10-year olds (M= 118.34, SD = 8.6), 12–23-year olds (M= 123.77, SD = 7.4), 15–16-year olds (M= 122.78, SD = 7.9) and 18–25-year olds (M= 121.30 SD = 10.6). Importantly, the different age groups did not differ in z-transformed IQ scores, $F(3,88) = 2.36$, $p = .075$.

Correlations were computed to determine whether IQ estimates were related to trust and reciprocity patterns. There was no significant correlation between z-transformed Raven SPM scores and the average percentage of trust ($r = .14$, $p = .17$) or reciprocity ($r = .17$, $p = .08$) decisions over all age groups or within each age group (all p's > .08). Nor were there significant relations between raw scores on the Raven SPM and trust or reciprocity (all p's > .1). Therefore these factors were not examined further.

2.3.2. Age differences in trust

Age groups differed in general trust percentage, $F(3,88) = 2.85$, $p < .04$, (see Fig. 2.3). Regression analysis across all participants with age as a covariate revealed a highly significant quadratic trend, $F(2,89) = 7.20$, $p = .006$, $r = .32$, and a mildly significant linear trend, $F(1,90) = 2.02$, $p < .037$, $r = .11$, between age and trust.

To investigate whether there was an effect of risk and benefit on trust decisions, we performed a repeated measures analysis of variance (ANOVA) with risk (high versus low) and benefit (high versus low) as a within-participants factor and age as a between-participants factor. For each participant we calculated the percentage of trust choices in each of the conditions[8]. In our initial analyses we also included gender as a between-participants factor. Because there were no significant effects of gender(all p's > .1) this factor was omitted from further analysis. Similarly we added IQ as a covariate to our ANOVA in order to control for differences in general intelligence. Doing so did not alter our results for the experimental manipulations and it was therefore also excluded in further analyses.

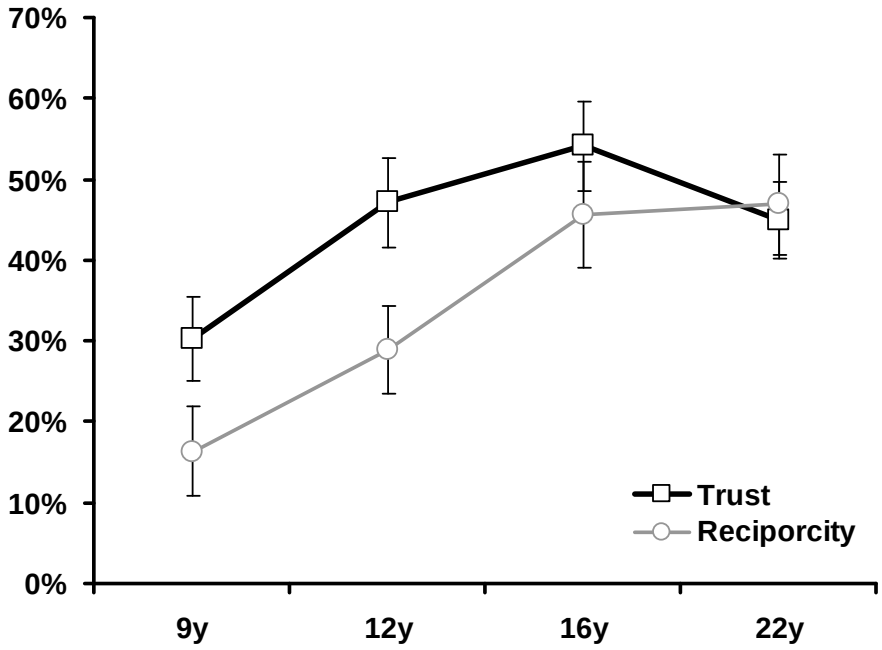


Figure 2.3: The mean percentage of trust and reciprocity decisions over the four experimental conditions for each of the four age groups (error bars represent S.E.).

[8] Because the percentage scores were not always normally distributed (confirmed with Shapiro–Wilk tests, with Lilliefors significance correction), we also analyzed the data using an arcsine transformation. These analyses yielded the same results as the ANOVAs on the untransformed data. To keep the statistics consistent with the behavioral data presented in the graphs, we present the analyses of untransformed data here.

As expected, high-risk trials resulted in fewer trust decisions than low-risk trials, $F(1,88) = 102.68$, $p < .001$. Performance did not differ significantly across age groups, $F(3,88) = 1.66$, $p = .18$. Although we observed no main effect of benefit, $F(3,88) = 2.34$, $p = .129$, we did observe a significant age by benefit interaction effect, $F(3,88) = 5.73$, $p < .001$. To further investigate the nature of the interaction with age we performed separate ANOVAs by age group. As seen in Fig. 2.4, 22-year olds trusted significantly more when the benefit for player 2 was high compared to low, $F(1,19) = 41.43$, $p < .001$, Bonferroni corrected. This difference was not significant for any of the younger age groups (all p's > .1).

In addition, when ANOVAs were performed for each age group separately, an interaction between risk and benefit was found in the youngest group, $F(1,22) = 16.14$, $p < .001$, Bonferroni corrected. Interestingly, 9-year-old children trusted less often when the risk was low and benefit was high, in contrast to all other age groups (Fig. 2.4). Note that no trust in the low-risk–high-benefit condition resulted in more money for player 1 than for player 2 (Fig. 2.2C). It is likely that the youngest age group trusted more often in this condition in order to avoid an outcome in which they got fewer coins than the other.

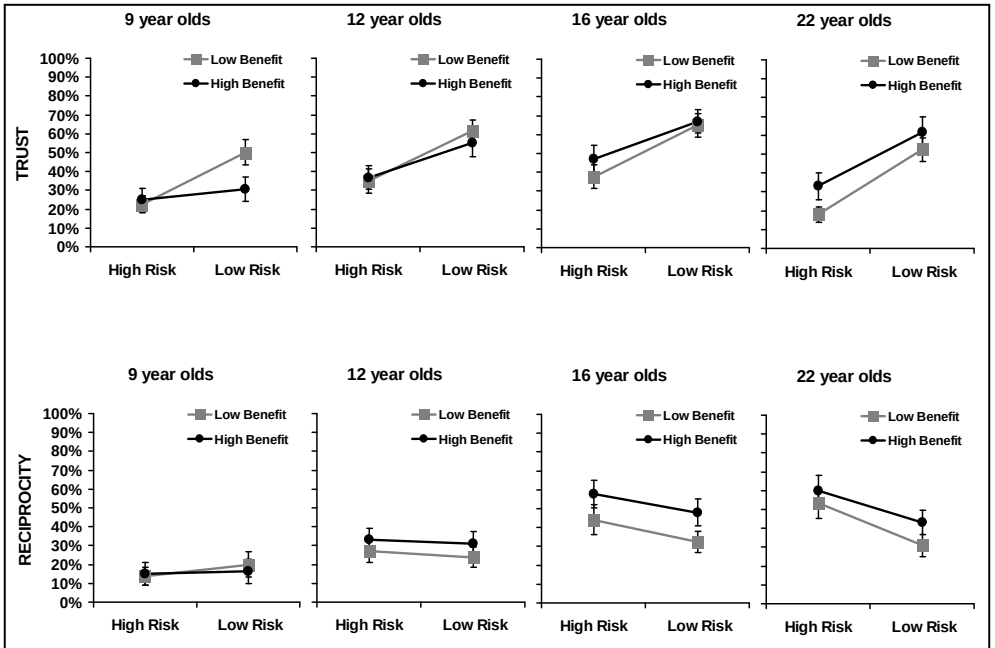


Figure 2.4: The mean percentage of trust and reciprocity decisions in each of the four experimental conditions for each of the four age groups (error bars represent S.E.).

To summarize, in addition to a general increase of trust with age we observed more trust decisions when there was (a) smaller risk for the participant and (b) higher benefit for the other player. The latter effect was only found for 22-year olds, showing that they differentiated more between high and low-benefit settings.

2.3.3. Age differences in reciprocity

Age groups differed in general reciprocity percentage, $F(3,88) = 5.69$, $p < .001$; see Fig. 2.3. Regression analysis across all participants with age as a covariate revealed a quadratic trend, $F(2,89) = 8.55$, $p < .001$, $r = .41$, and a linear trend, $F(1,90) = 16.33$, $p < .001$, $r = .37$, between age and reciprocity.

A repeated measures ANOVA was performed with risk (high versus low) and benefit (high versus low) as within-participants factors and age as between-participants factor for the percentage of reciprocal choices. Again, gender and IQ were dropped from these analyses because our initial analysis revealed no significant effect of gender or IQ on reciprocity (all p's > .1).

As expected, we found an effect of benefit on reciprocity, $F(1,88) = 24.14$, $p < .001$. Participants reciprocated more often when their benefit of being trusted was high rather than low. The effect of benefit on reciprocity differed between age groups: age x benefit $F(3,88) = 4.75$, $p < .005$. Post-hoc ANOVAs revealed that the effect of benefit was found in all age groups (all p's < .001, Bonferroni corrected), except for the 9-year olds who were insensitive to the benefit manipulation, $F(1,22) < 1$, $p = .50$.

In addition, we found a significant main effect of risk on reciprocity, $F(1,88) = 20.77$, $p < .001$. Participants were more willing to reciprocate when the risk taken by the other player was high. The risk effect was qualified by an age × risk interaction, $F(3,88) = 9.24$, $p < .001$. In general, participants reciprocated more often when the risk for the other player was high rather than low, but this difference was only significant for 16- and 22-year olds, $F(1,22) = 10.26$, $p < .005$, and $F(1,19) = 13.23$, $p < .005$, respectively, Bonferroni corrected. There was no risk effect for the two younger age groups (both p's > .1).

To summarize, in addition to a general increase of reciprocity with age we observed that increased benefit for the participant led to increased reciprocity and increased risk for the other player also led to more reciprocity. Increased benefit resulted in increased reciprocity in all age groups except for the 9-year olds, and increased risk resulted in increased reciprocity only for the 16- and 22-year olds.

2.3.4. Control condition

In the control condition we expected high levels of trust as player 1 and high levels of reciprocity as player 2, because these choices resulted in highest gain for both players. The results confirmed our expectations – all groups perform well above chance level – but there were also subtle differences across age groups. A univariate ANOVA with age as fixed factor and percentage of trust choices as dependent variable revealed a group difference, $F(3,88) = 7.95$, $p < .001$, showing that the youngest age group (9–10-year olds) made fewer trust decisions (75%) in their role as player 1 relative to the other age groups, confirmed by post hoc tests (12-, 16-, and 22-year olds; 92%, 95% and 98%, respectively), but they still performed well above chance level. A similar ANOVA for the percentage of reciprocal decisions by player 2 also resulted in significant age differences, $F(3,88) = 3.48$, $p < .02$. Post hoc tests revealed that the 9-year olds (88%) did not differ from the 12-year olds (96%), but the 9-year olds chose to reciprocate significantly less often than the two oldest age groups (16- and 22-year olds, 97% and 98%, respectively).

The lower trust scores by the youngest age group was unexpected, and therefore we reanalyzed the data including only the best performing half of the youngest group, based on a median split of the control scores. A comparison of the high-performing 9-year-old children and the other age groups no longer revealed age differences in the control condition: control trust, $F(3,72) = .72$, $p = .54$, and control reciprocity, $F(3,72) = .65$, $p = .58$. However, the effects on general trust and reciprocity, as well as those on risk and benefit, were not altered when the lower performing 9-year olds were removed from the analyses. This suggests that although there are developmental differences in performance on the control task, these are not related to differences on relevant task behavior.

2.3.5 Pay-off

Because age groups showed differences in types of decisions, they also obtained different amounts of coins during the game. There occurred an increase of total coins with age for the trustor and a decrease for trustee (Table 2.2). This is caused by the fact that trusting yields more coins than not trusting and exploiting yields more coins than reciprocating. Although the patterns of pay-offs differ, there are no significant differences between groups in total earnings, $F(3,88) = 1.67$, $p = .07$. Recall that the players knew that only the pay-off of a small number of rounds would be paid.

2.3.6 Time-on-task effects

Time-on-task effects were examined by dividing the task in three equal blocks. The original ANOVAs were repeated with blocks as an additional within-participants factor of three levels. All reported effects, for trustor as well as trustee, remained significant and did not result in any significant effects for block (all p's > .1). This result shows that participants did not change their patterns of behavior during the task.

Table 2.2 Average pay-off for each role for each age group.

	Player 1	Player 2
9 years	135.3 (7.0)	146.4 (9.3)
12 years	152.4 (7.5)	138.5 (8.5)
16 years	161.0 (8.1)	128.0 (9.0)
22 years	149.8 (8.8)	127.5 (9.8)

2.4 Discussion

This study had two main goals: (a) to develop a new version of the trust game that would allow us to examine the developmental trajectory of trust and reciprocity between late childhood and late adolescence, and (b) to examine the extent to which these processes are sensitive to the risk for the trustor and benefit of being trusted. To this end, the discussion is organized according to these main goals.

2.4.1 Developmental Trust Game

The Developmental Trust Game differs from most previous versions of the Trust Game in three important ways. First, the task was changed into a child-friendly game by making use of small amounts that were visually represented by coins, making sure the task had the same difficulty level for all age groups. Second, the computerized design made it possible to let the participants play multiple games against many different presumed players. This, in turn, made it possible to test each participant in each of the 5 conditions (experimental + control) multiple times, which allowed for robust within participant comparisons. Because we did not find any changes in behavior during the task, we are confident that our results are not due to time-on-task effects, which are possible side-effects of multiple rounds. Third, to our knowledge this is the first study in which participants played the role of trustor as well as trustee in an experiment with multiple trials. The performance of adults resembles the pattern typically seen in past work. That is, participants often chose to trust, suggesting that they expected others to reciprocate, even when decisions were anonymous. Also in line with previous results, adults often reciprocated even when doing so

was costly (Berg et al., 1995; McCabe et al., 2001). We made use of a fixed binary choice paradigm which allowed independent manipulation of risk and benefit in the DTG. As expected, both risk and benefit independently influenced the percentage of trust and reciprocal choices.

In accordance with past work (Malhotra, 2004), we found that adults were sensitive to risk manipulation as trustor and to benefit manipulation as trustee. As expected, participants were more willing to trust when risk was low and more willing to reciprocate when benefit was high. In addition, our study yielded two novel findings.

First, the benefit manipulation also influenced the decisions of the trustor. That is, 22-year olds trusted more often when the benefit for the other player was high rather than low. This increase in trust could be motivated by either altruistic inclination – participants care more about the welfare of the other with age – or by strategic intuition—they expect a higher change on reciprocity and therefore are more willing to trust in service of their own interest. Both explanations rely on more advanced forms of perspective-taking. In both cases the outcome for the other is valued, either intrinsically or instrumentally and integrated in the decision-making process.

Second, in 22-year olds risk manipulation also influenced the decisions of the trustee. In other words, the trustee was more willing to reciprocate when the risk for the trustor was high rather than low. This result suggests that the trustee appreciates the risk taken by the trustor and returns the favor by reciprocating. Playing both roles could have facilitated taking the perspective of the other player, which can be an explanation for the effects of risk and benefit present in the oldest age group but which are absent in a previous study with an adult population (Malhotra, 2004).

Together, the results of this study suggest that for adults trust and reciprocal decisions are not only dependent on their own outcome but also on the consequences for the other. The behavioral pattern of adults provides the framework for understanding developmental changes in trust and reciprocity.

2.4.2 Developmental changes in social decision-making

All age groups scored above chance level on the control task, indicating that the Developmental Trust Game is suitable for developmental research. However, the 9-year-old group scored lower than the adolescent groups. Given that they scored greatly above chance level and given the extensive training and the requirement of correct answers to assessment questions prior to the task, it seems unlikely that 9-year olds did not understand the task. A possible explanation is that 9-year olds did not want to wait for the ‘trust’ outcome and failed to show delay-of-gratification. This is consistent with several studies that show developmental differences on simple delay-of-gratification tasks that last until at least mid-adolescence (Green, Fry & Meyerson, 1994; Green, Myerson & Ostaszewski, 1999). However, future research is needed to investigate this hypothesis in more detail.

Consistent with earlier reports, there was an increase in both trust and reciprocity with age (Sutter & Kocher, 2007). Interestingly, although there were no age differences in overall earnings, children did earn more as the second player by not reciprocating as often as the older age groups. The decrease in earnings with age for the second player could be interpreted as a decrease in 'rational self-interest' behavior and potentially reflects an increase in showing socially desirable behavior. These results are important because they are consistent with prior reports suggesting that there is a general increase of prosocial behavior during adolescence that stabilizes between middle and late adolescence (Eisenberg et al., 1991, 1995; Schaffer, 1996).

In addition to these general developmental changes in trust and reciprocity, we also observed specific changes in trust and reciprocity related to the outcome manipulations as a function of age. First, there were important age related changes in trust decisions. Although all age groups were more willing to trust when the risk was low rather than high, there were age related changes in sensitivity to the benefit of the other player in trust decisions, as was evident for the 22-year olds. The possible motivations to take the consequences of the other player into account require a level of perspective-taking that appears to be present only in the oldest age group (late adolescence). In addition, although all age groups were more willing to trust when the risk was low rather than high, the 9-year-old children showed a slightly different pattern. They were more willing to trust when the risk was low and the benefit was low. This strategy might be explained by the fact that in the low-risk–high-benefit condition, the no-trust option resulted in a relatively higher outcome for player 1 than player 2, a situation which the youngest participants might wish to avoid. As such, this pattern suggests that they were also motivated by competitive motives. This is consistent with previous literature showing that competitive social value orientation – preference for increasing relative gain over others – decreased during adolescence (Van Lange et al., 1997) and another study by Fehr, Bernhard and Rockenbach (2008) showing that children are competitively oriented in social situations.

Second, there were also age related changes in reciprocal decisions. The effect of benefit on reciprocity was present in early adolescence, indicating that in this period basic reciprocity emerges. In contrast, 9-year-old children do not yet show this type of behavior. From middle adolescence onwards, amore elaborate form of reciprocal behavior appeared. At this point participants also chose to reciprocate the risk taken by the other player (trustor).

A comparison of age differences in sensitivity to risk and benefit for trust and reciprocity suggests that, besides a general increase of prosocial behavior, considering the outcomes for the other becomes more important in social decision-making during adolescence. Here this type of perspective-taking was examined in the context of prosocial behavior, but it should be noted that increased perspective-taking ability can also be used for strategic or anti-social purposes, such as lying and cheating (Rotenberg, 1991; Beate & Frith, 1992).

To our knowledge, there are no experimental studies that have investigated both the development of on-line prosocial behavior and development of perspective-taking during adolescence. Prior studies have suggested that there are subtle developmental changes on experimental measures of perspective-taking during adolescence (Choudhury, Blakemore, & Charman, 2006; Duhmontheil, Apperly & Balkemore, 2009), but these studies did not examine perspective-taking in a social context. Our data also suggest that later in adolescence there is no general increase in prosocial behavior but rather a sophistication of prosocial behavior. Although trust and reciprocal behavior were at a stable level at mid-adolescence, there were still changes in the effect of the outcome manipulations until late adolescence. Thus, with age, prosocial behavior becomes more context dependent, leading to more prosocial behavior in one context (e.g. a high-risk and high-benefit situation) but less in another (a low-risk and low-benefit situation).

Finally, our current results do not speak to the issue of a presumed relation between behavioral measures of taking into account the intentions of and consequences for the other and other direct measures of perspective-taking. It would therefore be interesting for future studies to include additional measures of perspective-taking. One way to shed more light on this research question would be to ask participants to think aloud while performing these tasks. Furthermore, it would be interesting to extend the present research involving a generalized other by studying interaction with specific others such as peers or parents.

3. What motivates repayment? Neural correlates of reciprocity in the Trust Game

Reciprocity of trust is important for social interaction and depends on individual differences in social value orientation (SVO).Here, we examined the neural correlates of reciprocity by manipulating two factors that influence reciprocal behavior: (1) the risk that the trustor took when trusting and (2) the benefit for the trustee when being trusted. FMRI results showed that anterior Medial Prefrontal Frontal Cortex (aMPFC) was more active when participants defected relative to when participants reciprocated, but was not sensitive to manipulations of risk and benefit or individual differences in SVO. However, activation in the right temporal parietal-junction (rTPJ), bilateral anterior insula and anterior cingulate cortex (ACC) was modulated by individual differences in SVO. In addition, these regions were differentially sensitive to manipulations of risk for the trustor when reciprocating. In contrast, the ACC and the right dorsolateral prefrontal cortex were sensitive to the benefit for the trustee when reciprocating. Together, the results of this study provide more insight in how several brain regions work together when individuals reciprocate trust, by showing how these regions are differentially sensitive to reciprocity motives and perspective-taking.

3.1 Introduction

One of the key components of human social interaction is cooperation or the exchange of favor or goods between individuals for the attainment of mutual benefit. Cooperation depends to a large extent on trust and reciprocity. Trust is required because cooperative exchanges are often separated in time, whereas reciprocity, or the repayment of what others have provided us, is thought to be important for the maintenance of social relationships. That is, if favors are not returned relationships may be short-lived (Lahno, 1995).

Both the trustor and the trustee may obtain higher outcomes when trust is given relative to when no trust is given. However, trusting also involves a component of risk, because the trustor may attain higher personal benefit when not reciprocating. Consequently, trusting may result in a smaller outcome for

the trustor relative to when the trustor would not have trusted (Rousseau et al., 1998). Thus, the decision to trust another party involves risk for the trustor and the decision to reciprocate trust depend on the offset between maximizing personal outcomes relative to the appreciation of the trust that was given (i.e. repayment). This study will focus on different motives involved in reciprocal behavior.

Researchers have demonstrated that even for single anonymous transactions, individuals often reciprocate trust even when this leads to a smaller personal monetary outcome (Berg et al., 1995; McCabe et al., 2001). It has therefore been suggested that our motivation to reciprocate trust is not only guided by goals to maximize personal outcomes, but also by other-regarding preferences (Falk and Fischbacher, 2006; Fehr and Camerer, 2007; Fehr and Gintis, 2007; Van Lange, 1999). According to these studies, the decision to reciprocate is dependent on evaluating consequences for both self and others. Importantly, reciprocal behavior is dependent on individual differences in social value orientation (SVO), the general tendency of individuals to value the outcome of others (McClintock and Allison, 1989; De Dreu and Van Lange, 1995; Van Lange et al., 1997). Furthermore, decisions to reciprocate trust are not only motivated by outcome considerations but also involve considerations of the intentions of others, such as the risk that the trusting party took when trusting or the benefit for the trusted party when being trusted. Therefore, these decisions are thought to be dependent on our ability to take the perspectives of others.

Neuroimaging studies in combination with game theoretical paradigms have investigated the neural correlates of the cognitive processes involved in cooperation and reciprocal exchange (e.g. King-Casas et al., 2005; Krueger et al., 2007; McCabe et al., 2001; Rilling et al., 2002). Several of these neuroimaging studies have reported activation in the anterior medial prefrontal cortex (aMPFC) when participants are involved in interactions with another person relative to a computer (McCabe et al., 2001; Rilling et al., 2004), and when participants decide to trust relative to when they decide not to trust (McCabe et al., 2001; Delgado et al., 2005; King-Casas et al., 2005; Krueger et al., 2007; Baumgartner et al., 2008). Prior neuroimaging studies have considered the aMPFC together with the temporal-parietal-junction (TPJ) to be important for mentalizing and theory-of-mind. For example, neuroimaging studies have demonstrated that aMPFC and TPJ are active during theory-of-mind tasks, such as tasks that require participants to infer mental states of characters in stories (Fletcher et al., 1995) and cartoons (Gallagher et al., 2002) or while watching animations (Castelli et al., 2000). In addition, prior studies have suggested that in a social context the aMPFC is involved in evaluating the

mental content of others in relation to the self (Amodio and Frith, 2006), whereas the TPJ is thought to be important for redirecting or focusing attention on the other (Mitchell, 2008). However, the mentalizing requirements during these theory-of-mind tasks are complex, and therefore it is difficult to dissociate the putative roles of the aMPFC and TPJ in social interaction (Hampton et al., 2008). Therefore, it remains to be determined how activation in aMPFC and TPJ can be associated with the different processes, which may underlie reciprocal exchange.

Besides the aMPFC and TPJ, neuroimaging studies of social decision-making have also suggested that brain regions that are associated with reward processing and arousal can mark social interactions as positive or aversive. For example, one neuroimaging study demonstrated that activation in the ventral striatum correlates positively with cooperation choices in a Prisoners Dilemma Game (Rilling et al., 2004). Two other neuroimaging studies showed that unfair treatment by a partner in the Ultimatum Game results in increased activation in the insula (Sanfey et al., 2003; Tabibnia et al., 2008), and this region has also been engaged during unreciprocated trust (Rilling et al., 2008). A recent study, which examined iterated two-person trust exchanges, demonstrated that the insula is more active for low relative to high levels of reciprocity. This finding was explained by suggesting a role of the insula in signaling personal norm violations (King-Casas et al., 2008). Thus, the ventral striatum and the insula seem to be involved in the pleasant and unpleasant aspects of social interactions, which may explain how lower level affective processes can result in encouragement or discouragement of social behavior (Sanfey, 2007). However, even though this pattern of activity is consistent over a wide range of social interactions paradigms, it has not been shown how these regions are associated with the choice and motivation to reciprocate.

Finally, the anterior cingulate cortex (ACC) and the right dorsolateral prefrontal cortex (rDLPFC) are typically engaged when individuals make decisions in which there is conflict between social norms and personal interest (Sanfey et al., 2003; Spitzer et al., 2007) or when individuals make decisions that may be counter to their own response tendencies (Rilling et al., 2002, 2007). In addition, transcranial magnetic stimulation of the right DLPFC lead to an increase of accepting unfair offers in the Ultimatum Game (Knoch et al., 2006). These control-related structures may therefore be involved in overriding self-oriented impulses.

Neuroimaging methods may allow us to examine the possible dissociations between different processes that underlie an individual's decision to reciprocate. Indeed, the review of prior neuroimaging studies suggests that the brain regions, which have been reported in social interaction studies, may indeed contribute in

different ways to different motives for reciprocity. However, to date, most neuroimaging studies of social interaction have examined the neural correlates of different types of choices (e.g. reciprocate vs defect) but have not attempted to dissociate between processes that may underlie the decision to reciprocate or defect, such as the risk that the trusting party took or the benefit the trusted party gained by being trusted. Therefore, the question remains how the brain regions, which have previously been associated with lower-level cognitive and affective processes and have been suggested to be involved in social interaction, are differentially involved in reciprocal behavior. This question can be addressed by investigating how these brain regions are differentially sensitive to the putative motives for reciprocity, which have been outlined above. In this study, we will manipulate the risk for the trustor and the benefit for the trustee, and we will examine the effects of these manipulations on the neural correlates of reciprocal behavior under these conditions. Thus, the goal of the current study was to determine whether the appreciation of different motives for reciprocity can be dissociated on a neural level by manipulating the risk that the trustor took when trusting and the benefit for the trustee when being trusted.

Participants played several one-shot rounds of the Trust Game, in which they had to make the decision whether or not to reciprocate trust given by another individual (Berg et al., 1995). In the Trust Game, two anonymous players are involved in dividing a certain amount of money. The first player (trustor) has two options. One option is to divide the money according to a predetermined scheme (e.g. eight for first player and seven for second player; see Figure 3.1 A), the other option is to trust the second player (trustee) and to give him/her the choice to divide the money. The latter option potentially leads to a higher pay-off for both players. If trusted, the second player has two options: (1) reciprocate the trust given by the first player (e.g. 11 for first player and 10 for second player) or (2) defect and maximize personal gains (e.g. 5 for first player and 17 for second player). All participants were assigned to the role of the second player and always had two fixed choices. This design allowed us to (a) concentrate on the decision to reciprocate or not and (b) systematically vary the main variables of interest: the risk for the trustor and the benefit for the trustee.

We predicted that the extent to which second players are motivated to reciprocate depends on the risk that the first player has taken (i.e. the amount of money the first player can lose by trusting) and the benefit that the second player receives when being trusted (i.e. the amount of money that the second mover receives when trusted relative to not being trusted) (Pillutla et al., 2003; Malhotra, 2004; van den Bos et al., 2010). More specifically, we expected that

participants were more motivated to reciprocate when either the risk or the benefit was high rather than low.

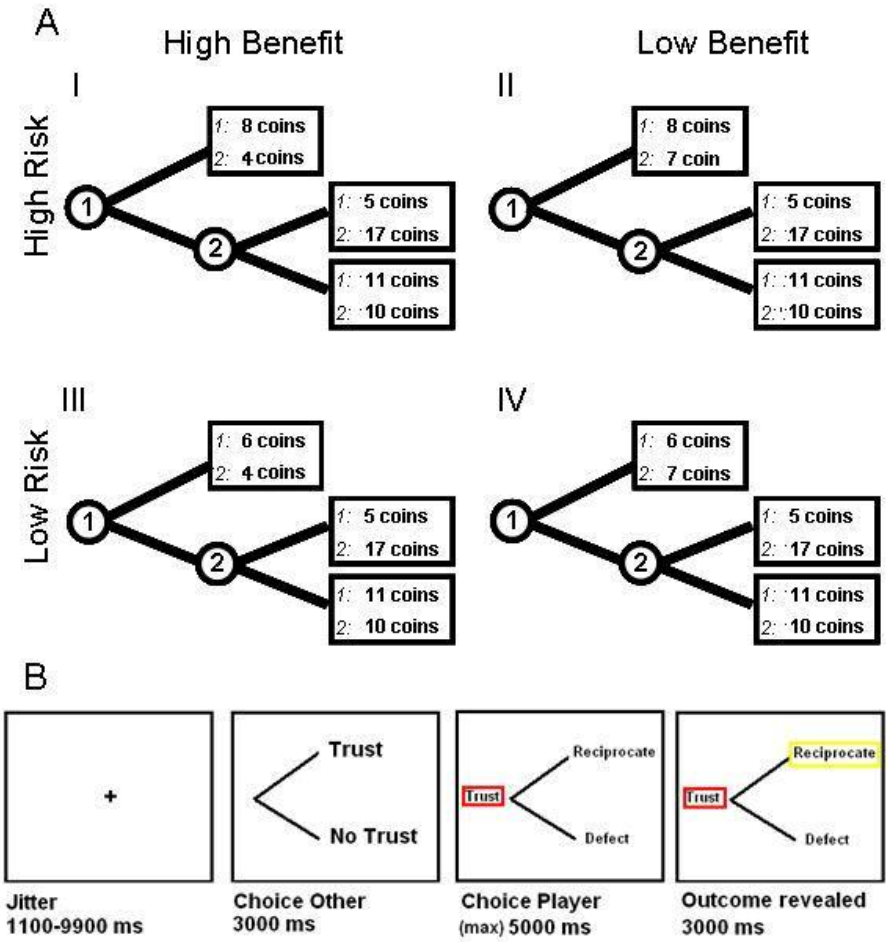


Figure 3.1: A) Presentation of task conditions. In four different conditions the risk that the trustor took when trusting and the benefit that the trustee received when being trusted were manipulated independently (Malhotra, 2004). **B)** Timing of the events in the scanner task in milliseconds.

We hypothesized that regions that are involved in mentalizing would be modulated by both risk and benefit manipulations. However, we expected that the type of perspective-taking would be associated with distinct neural correlates. In particular, we posited that regions that are important for taking the perspective of the other would be especially sensitive to the risk manipulation because the risk manipulation requires participants to take into account the outcomes of the other (first) player. Thus, the risk manipulation focused on neural correlates of mentalizing about how the different outcomes affect the

first player. In contrast, we posited that regions, which are associated with self-referential thought, would be sensitive to the benefit manipulation, because the benefit manipulation involves taking into account the second player's own increased outcome in case of trust. Thus, the benefit manipulation focused on neural correlates of mentalizing about the cooperative intentions of the first player, which benefits the second player.

We predicted that aMPFC and TPJ would exhibit a pattern consistent with their suggested roles in perspective-taking. In particular, we expected that the risk manipulation, motivating participants to take the perspective of the outcomes for the other, would result in a shift in attention from self to the other and thus would be associated with changes in TPJ activity (Lamm et al., 2007). On the other hand, we expected that the aMPFC would be more engaged by the benefit manipulation, because this manipulation motivated the participants to consider their own outcomes and the cooperative intentions of others (McCabe et al., 2001; Gallagher et al., 2002; Hampton et al., 2008).

We expected that the ACC and rDLPFC would also be sensitive to risk and benefit manipulations and would exhibit a pattern consistent with a role in overcoming selfish impulses (Rilling et al., 2002, 2007; Knoch et al., 2006). Therefore, we expected that these regions were most engaged when the participants reciprocated in situations where the incentive to reciprocate was low (low-benefit condition).Finally, we predicted that the insula would be sensitive to situations, which involved violations of one's own behavioral norms (Montague and Lorenz, 2007; King-Casas et al.,2008). Therefore, we expected a pattern of activation partly overlapping with activation observed in ACC and rDLPFC. In the insula, we expected increased activation when reciprocating in both low-benefit and low-risk conditions.

Finally, we expected that the need and/or engagement of the affective and control regions would also be dependent on the internal motivations to reciprocate. As such, the individual differences in reciprocal behavior in the current task were related to scores on the SVO questionnaire (van Lange, 1999), which is a personality variable that indicates how people evaluate outcomes for themselves and others. This questionnaire has shown significant external validity in a variety of settings (McClintock and Allison, 1989; De Dreu and van Lange, 1995; van Lange et al., 1997). Prosocial personalities were expected to reciprocate more often than the proself personalities (Kramer et al., 1986). We posited that the activity in regions, which are associated with affective processes, would also correlate with individual differences in SVO. The insula and striatum were predicted to be sensitive to individual predispositions to reciprocate or defect reflecting differences in social norms and preferences. By the same token, we expected that prosocial participants would show less activity

in the control network (DLPFC, ACC) when reciprocating than the proself individuals and that proself participants would show more activation in the control network when reciprocating.

3.2 Methods

3.2.1 Participants

Twenty-two healthy right-handed paid volunteers (11 female, 11 male; age 18–22, M = 19.7, SD. = 1.3) participated in the fMRI experiment. Four of the participants were excluded from the analysis, because there were missing cases in one or more conditions (i.e. only reciprocal choices or only defect choices, see supplementary data). Subsequent fMRI analyses were based on the remaining 18 participants (nine female, nine male; age 18–22, M = 19.7, SD. = 1.4). All participants reported normal or corrected-to-normal vision and an absence of neurological or psychiatric impairments. All participants gave informed consent for the study, and all procedures were approved by the Leiden University Department of Psychology and the medical ethical committee of the Leiden University Medical Center. In accordance with Leiden University Medical Center policy, all anatomical scans were reviewed by the radiology department following each scan. No anomalous findings were reported. Standard intelligence scores were obtained from each participant using the Raven's Progressive Matrices test. All participants had average or above average IQ scores (M = 116.12, SE = 1.98).

3.2.2 Task

Trust Game. During the fixed choice Trust Game (Berg et al., 1995; Malhotra, 2004), participants were instructed that in an earlier phase of the study, other individuals had been assigned the roles of first player and that they would complete the second phase of the study in the role of second player. They were instructed that they were not playing directly with first players, but that they played with the implementation of answers of first players which were gathered in the previous part of the experiment. They were explained that their decisions would have consequences for the first player and that the payment of all participants would take place after completion of the experiment.

Each round, participants were paired with a different, anonymous player to exclude reputation effects or strategy use, and the other players were matched for gender. For those trials where the first players had decided to trust, the participant was presented with two options: reciprocate or defect. If the participant decided to defect, the participant would maximize his/her own gains

and the first player would receive less money than in the no-trust option. In case the participant reciprocated, the money was shared almost equally and both players received more money compared to the no-trust option, but the second player received less money compared to when he/she would have defected (see Figure 3.1A). Participants were instructed that at the end of the experiment the computer would randomly select the outcome of five trials, and the sum of these trials would determine the pay-off for the participant and for the first players. Consequently, their decisions had implications for both their own pay-off as well as that of the other players.

Each trial started with a 3 s display of the choice alternative for the first player, followed by the trust or no-trust decision of the first player. For those trials on which the first player chose not to trust, the no-trust decision was visually presented for 3 s. For those trials on which the first player chose to trust, the defect and reciprocate options were presented, and participants were instructed to make their decision by pressing the middle or index finger of the right hand. Participants were instructed to respond within a 5 s window (see Figure 3.1B). The 5 s decision-display was followed by a 3 s display of their choice.

Risk for the trustor (high vs low) and benefit for the trustee (high vs low) were manipulated separately (Malhotra, 2004) (see Figure 3.1A). The risk manipulation determined the risk for the first player. In the high-risk condition, the first player could lose a large amount of money by trusting the participant in case the second player chose to defect. In contrast, in the low-risk condition, the first player could lose only a small amount of money by trusting the second player. The benefit manipulation determined the benefit for the second player when being trusted. In the low-benefit condition, the difference between money gained by player 2 when being trusted relative to not being trusted was small. In contrast, in the high-benefit condition, the increase of money for the second player by being trusted was large. The risk and benefit manipulations were based on the Malhotra (2004) paradigm.

The computer played a fixed strategy that was based on behavior of participants in previous studies (van den Bos, et al., 2010). In total, the task consisted of 43 high risk-high benefit trials (25 trusted, 18 not-trusted), 44 high risk-low benefit trials (23 trusted, 21 not trusted), 48 low risk-high benefit trials (35 trusted, 13 not-trusted) and 53 low risk-high benefit trials (42 trusted, 11 not-trusted). Consequently, for each participant, the task consisted of 188 rounds in total, with 125 trusted trials, which required a decision from the participant. The trials were divided over five blocks, each block lasted ~8.5 min. The trials were presented in pseudo-random order with a jittered

interstimulus interval (min. = 1.1s, max. = 9.9s, mean = 3.37s) optimized with OptSeq2 [surfer.nmr.mgh.harvard.edu/optseq/, developed by Dale (1999)].

Social Value Orientation. All participants completed the SVO questionnaire. The SVO is a brief measure of allocation choices between self and other and has shown significant external validity in a variety of settings. The questionnaire consists of nine tables or 'decomposed games' [for more details, see van Lange (1999)]. In these decomposed games, the participant determines the outcome for both himself and a hypothetical other.

The three different decompositions correspond to three different types of SVOs: (1) a cooperative orientation, reflecting a preference for joint outcomes, (2) an individualistic orientation, reflecting a preference for own outcomes and (3) a competitive orientation, reflecting a preference for a large positive difference between own and other outcomes. When participants make six or more consistent choices in nine games, they are classified as belonging to one of three types of SVO: cooperative, individualistic or competitive. In prior studies, cooperative participants have been categorized as a 'prosocial' group, and individualistic and competitive participants have been categorized as a 'proself' group. The reason for the latter categorization is based on the observation that both individualistic and competitive individuals value outcomes for self higher than outcomes for others (van Lange, 1999).

Task Procedure. Prior to the experiment, participants received oral instructions and completed a practice session (20 trials). The stimuli and timing of the practice sessions were the same as in the fMRI experiment. The Raven SPM and SVO questionnaire (Van Lange, 1999) were administered after the scanning session. The total duration of the experiment was ~2 h.

MRI Procedure. Data were acquired using a 3.0T Philips Achieva scanner at the Leiden University Medical Center. Stimuli were projected onto a screen located at the head of the scanner bore and viewed by participants by means of a mirror mounted to the head coil assembly. First, a localizer scan was obtained for each participant. Subsequently, T2*-weighted EPI (TR = 2.2 s, TE = 30 ms, 80x80 matrix, FOV = 220, 352.75-mm transverse slices with 0.28mm gap) were obtained during five functional runs of 232 volumes each. The first two scans were discarded to allow for equilibration of T1 saturation effects. A high resolution T1-weighted anatomical scan and a high resolution T2-weighted matched-bandwidth high-resolution anatomical scan (same slice prescription as EPI) were obtained from each participant after the functional runs. Stimulus

presentation and the timing of all stimuli and response events were acquired using E-Prime software.

fMRI Data Analysis. Data were preprocessed using SPM2 (Wellcome Department of Cognitive Neurology, London). The functional time series were realigned to compensate for small head movements. Translational movement parameters never exceeded 1 voxel (< 3 mm) in any direction for any subject or scan. Functional volumes were spatially smoothed using a 6mm full-width half-maximum Gaussian kernel. Functional volumes were spatially normalized to EPI templates. The normalization algorithm used a 12-parameter affine transformation together with a nonlinear transformation involving cosine basis functions and resampled the volumes to 3mm cubic voxels. The MNI305 template was used for visualization and all results are reported in the MNI305 stereotaxic space (Cosoco et al., 1997), an approximation of Talairach space (Talairach and Tourneaux, 1988).

Statistical analyses were performed on individual participants' data using the general linear model in SPM2. The fMRI time series data were modeled by a series of events convolved with a canonical hemodynamic response function (HRF). The start of the first player's choice display and the start of the second player's choice display (only for trust trials) of each trial were modeled as zero-duration events. The second player's choice display condition was divided in trust and no-trust choices and the trust choices were divided into reciprocate and defect decisions. Finally, those choices were further divided in four experimental conditions (high vs low risk¬ high vs low benefit). These trial functions were used as covariates in a general linear model, along with a basic set of cosine functions that highpass filtered the data and a covariate for run effects. The least-squares parameter estimates of height of the best-fitting canonical HRF for each condition were used in pairwise contrasts. The resulting contrast images, computed on a subject-by-subject basis, were submitted to group analyses. At the group level, contrasts between conditions were computed by performing one-tailed t-tests on these images, treating participants as a random effect. Mean reciprocity levels were used in regression analyses to test for brain–behavior relations. We applied AlphaSim (Ward, 2000) to calculate the appropriate threshold significance level and cluster size. A significance threshold of $p < 0.05$, corrected for multiple comparisons was calculated by performing 10 000 Monte Carlo simulations in AlphaSim resulting in an uncorrected threshold of $p < 0.001$, requiring a minimum of 12 voxels in a cluster.

Region-of-Interest (ROI) Analyses. ROI analyses were performed to further characterize sensitivity to risk and benefit manipulations. Averaging the signal across voxels, as is done in ROI analyses, captures the central tendency and tends to reduce uncorrelated variance. Thus, ROI analyses have greater power than whole-brain statistical contrasts to detect effects that are present across a set of voxels. ROI analyses were performed with the Marsbar toolbox in SPM2 (Brett et al., 2002; http://marsbar.sourceforce.net/).

The contrast used to generate functional ROIs based on a priori hypotheses was that of all choices > fixation, unless otherwise specified in the text. Functional maps were masked with anatomical masks from the Marsbar toolbox. For all ROI analyses, effects were considered significant at an α of 0.008, based on Bonferonni correction for multiple comparisons (p = 0.05/0.06 ROIs (aMPFC, rTPJ, rDLPFC, ACC, anterior insula and ventral striatum), unless reported otherwise. For each ROI, the center of mass is reported.

3.3 Results

3.3.1 Behavioral data

Trust Game. On average, participants reciprocated half of the trials (M = 51%), but there were large individual differences in behavior (SD = 18%, min. = 22%, max. = 78% see supplementary results). To investigate whether there were effects of the risk and benefit manipulations on reciprocity decisions, we performed a repeated measures ANOVA with risk (high vs. low) and benefit (high vs. low) as within-subject factors. As expected, high risk for the first player resulted in more reciprocal choices (59%) than low risk for the first player (43%) (main effect risk, $F(1,18) = 26.85$, $p < 0.001$) and high benefit for the second player resulted in more reciprocal choices (61%) than low benefit for the second player (40%) (main effect benefit $F(1,18) = 22.03$, $p < 0.001$). In addition, there was a significant risk x benefit interaction [$F(1,18) = 9.92$, $p < 0.01$]. This interaction demonstrated that the difference between high- and low benefit reciprocal choices was larger for low risk trials (high benefit: 58%, low benefit: 27%) than for high-risk trials (high benefit: 64%, low benefit: 53%). Thus, when the risk to trust was high for the first player, participants focused less on their own benefit when deciding to reciprocate. Finally, there were no differences in mean reaction times for defect (M = 1.77 s, SE = 0.13) vs. reciprocate (M = 1.76 s, SE = .12) choices [$t(21) = 0.044$, P = 0.96].

Social Value Orientation. Classification of participants by SVO (Van Lange, 1999) resulted in 8 proself and 10 prosocial-oriented individuals. The SVO was a strong predictor of reciprocal behavior in the Trust Game as administered in

the scanner session. A t-test for reciprocity level demonstrated that prosocial individuals reciprocated significantly more (M = 62%, SD = 11%) than proself individuals [M = 39%, s.d. = 10%; t(1,16) = 3.72, p < 0.002]. When reciprocity levels in the Trust Game were divided based on a median split analysis, the low-reciprocity group consisted of all eight proself classified participants and one prosocial classified participant. The high-reciprocity group consisted of only prosocial classified participants. Thus, performance in the current version of the Trust Game had high external validity as demonstrated by a high correlation with SVO.

3.3.2 fMRI data

Whole Brain Results - Main effects. To examine the neural correlates of reciprocity, we examined neural activity for reciprocate and defect choices for those trials on which the participant was trusted. The comparison of defect choices > reciprocate choices revealed activity in the aMPFC (BA 32; Figure 3.2A, Table 3.1) and the primary visual cortex (MNI 6,-93, 12), whereas the opposite contrast (reciprocate > defect) resulted in significant activation only in primary visual cortex (MNI 9, -63, 12). It should be noted that defect and reciprocate alternatives were always displayed on the same location of the screen, which may explain the consistent activation in the visual areas for the separate contrasts.

Regression analysis. The second set of contrasts aimed at revealing individual differences in neural activation by adding average reciprocity level as a predictor variable to a regression analysis. This analysis revealed a positive correlation between levels of reciprocity and BOLD activity for defect > reciprocate choices in the dorsal ACC, bilateral anterior insula, right TPJ (rTPJ) and precuneus (Figure 3.2B). Those individuals who generally showed prosocial behavior by reciprocating more often also showed increased activation in these areas when defecting. In contrast, those individuals who reciprocated less often showed more activation in these areas when reciprocating (see also supplementary results). Thus, these areas were sensitive to the less frequently chosen alternative, regardless of whether the less frequent alternative was to reciprocate or to defect.

There were no regions that showed a negative correlation between reciprocity and BOLD activation for defect > reciprocate at a $p < 0.001$ threshold. However, lowering the threshold to an uncorrected threshold of $p < 0.05$ revealed a negative correlation between reciprocity and the defect > reciprocate contrast in the ventral striatum. Here, individuals who reciprocated more often showed increased activation when reciprocating, and individuals

who reciprocated less often showed less activation when reciprocating (see supplementary results for performance correlations).

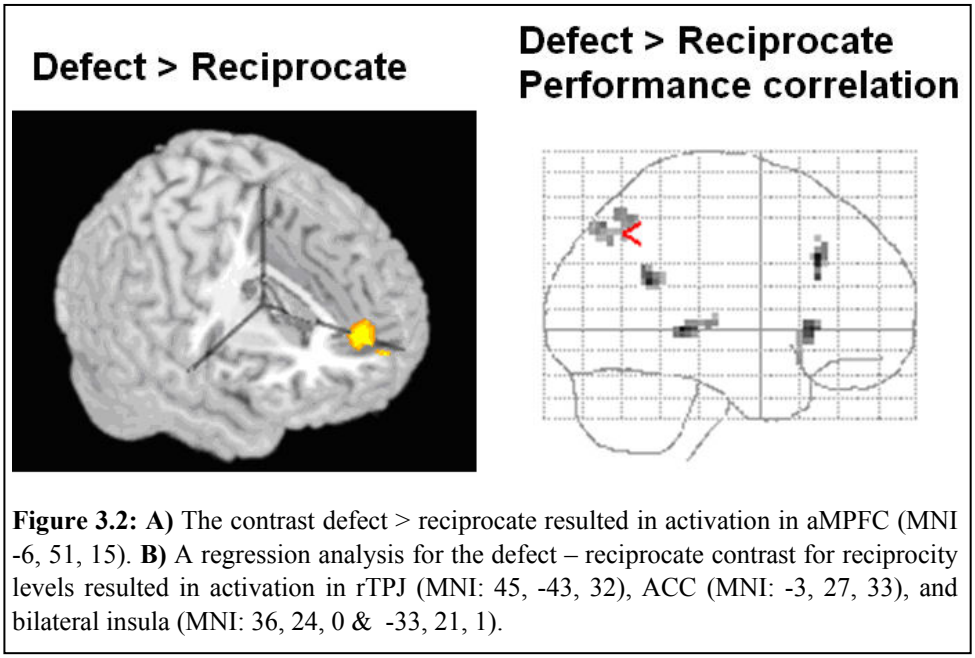


Figure 3.2: A) The contrast defect > reciprocate resulted in activation in aMPFC (MNI -6, 51, 15). **B)** A regression analysis for the defect – reciprocate contrast for reciprocity levels resulted in activation in rTPJ (MNI: 45, -43, 32), ACC (MNI: -3, 27, 33), and bilateral insula (MNI: 36, 24, 0 & -33, 21, 1).

Table 3.1: Brain Regions revealed by whole brain contrasts and regressions analysis

Anatomical region	L/R	Volume (mm)	Z	MNI coordinates		
				x	y	z
Main effect of Choice						
Defect > Reciprocate						
Paracingulate cortex, VMPFC	L	666	5.84	-6	51	15
Visual Cortex	L/R	1006	6.06	6	-93	12
Reciprocate > Defect						
Visual Cortex	L/R	720	4.43	9	-63	3
Regression Defect > Reciprocity			Z			
Positive corr. avg. reciprocity						
Anterior Cingulate Cortex	L/R	917	4.10	-3	27	33
Anterior Insula	R	371	4.06	36	24	0
Anterior Insula	L	286	3.97	-33	21	1
Temporal Parietal Junction	R	862	4.06	45	-43	32
Precuneus	L	423	3.32	-24	-72	45
Thalamus	R	223	3.91	6	-30	0
Negative corr. avg. reciprocity**						
Ventral Striatum	R	171	1.63	14	12	-5

MNI coordinators for main effects, peak voxels reported at $p < .001$, at least 10 contiguous voxels.** peak voxel reported at $p < .05$

ROI analyses. ROI analyses were performed to further characterize sensitivity to risk and benefit manipulations. For these analyses, we focused on six a priori defined regions: aMPFC, rTPJ, rDLPFC, ACC, anterior insula and ventral striatum. rDLPFC, ACC and ventral striatum were derived from the all choices > fixation contrast. Not all regions were revealed by this contrast; therefore, aMPFC was selected based on the defect > reciprocate contrast, and the right TPJ and right insula were derived from the regression analyses.

Because our hypotheses concerned the modulations of the neural correlates of reciprocal choices, we analyzed the effects of the risk and benefit manipulations for reciprocal choices. We used ANOVA to analyze BOLD differences that accompanied the choices to reciprocate and to characterize possible interactions with risk and benefit manipulations.These analyses revealed main effects of benefit in the ACC [$F(1, 17) = 5.46$, $p = 0.01$, Figure 3.3A] and the rDLPFC [$F(1, 17) = 9.98$, $p < 0.003$; Figure 3.3B]. These analyses demonstrated that there was greater activation in both the ACC and the rDLPFC when participants chose to reciprocate when the benefit for themselves was low relative to when the benefit for themselves was high. Thus, ACC and rDLPFC were more active when participants decided to reciprocate, even though the benefit of being trusted was low.

There was also a main effect of risk in the right TPJ [$F(1, 17) = 6.43$, $P = 0.01$, Figure 3.4A]. In this region, more activation was observed for reciprocate choices when the risk for the first player was high relative to when the risk for the first player was low. Finally, there was a main effect of risk in the right insula [$F(1, 17) = 8.80$, $P < 0.005$, Figure 3.4B], but opposite to the risk effect in the rTPJ, this region was more active when participants chose to reciprocate when the risk for the first player was low relative to when the risk for the first player was high. Thus, rTPJ was more active when participants decided to reciprocate and repaid the risk that was taken by the first player. In contrast, the right insula was more active when participants reciprocated despite the low need for repayment. Finally, there were no effects of risk or benefit for the aMPFC or the striatum.

Frequency Effects. Because the changes in activation can be influenced by frequency effects, we correlated activation in the ROIs with the frequency of different types of behavior to test whether the reported effects of risk and benefit can be explained by frequency differences. In addition, we added the frequency of behavior as a covariate of interest in ANCOVAs. Together, these effects showed that the risk and benefit effects were not correlated with frequency of choices, except for neural activation in the insula (see

supplementary data). That is, activation in the insula was highest for the least frequently occurring choices.

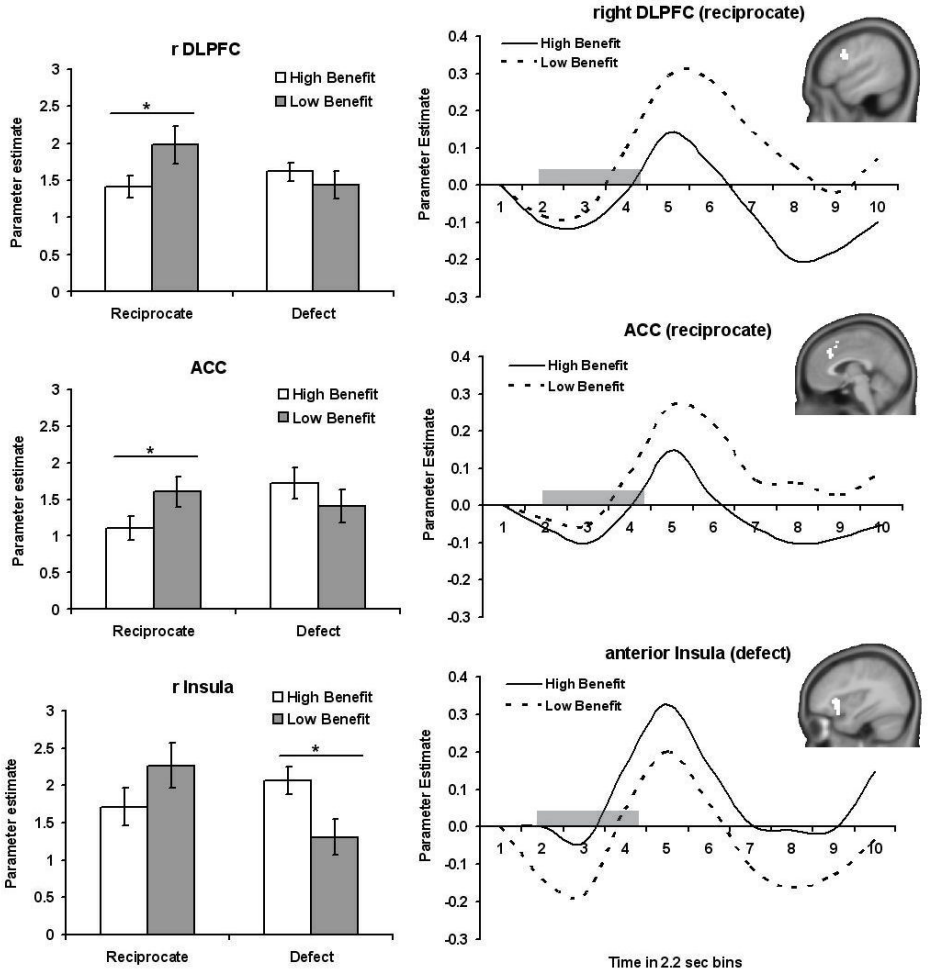


Figure 3.3: ROI parameter estimates and time series for regions that were sensitive to the benefit manipulation (error bars represent standard error). ACC (MNI: -3, 27, 33) and right DLPFC (MNI 51, 18, 30) were more active for reciprocate choices where the benefit of being trusted was low relative to high. The time-series plots show the data collapsed over conditions, gray area represent decision window of participant.

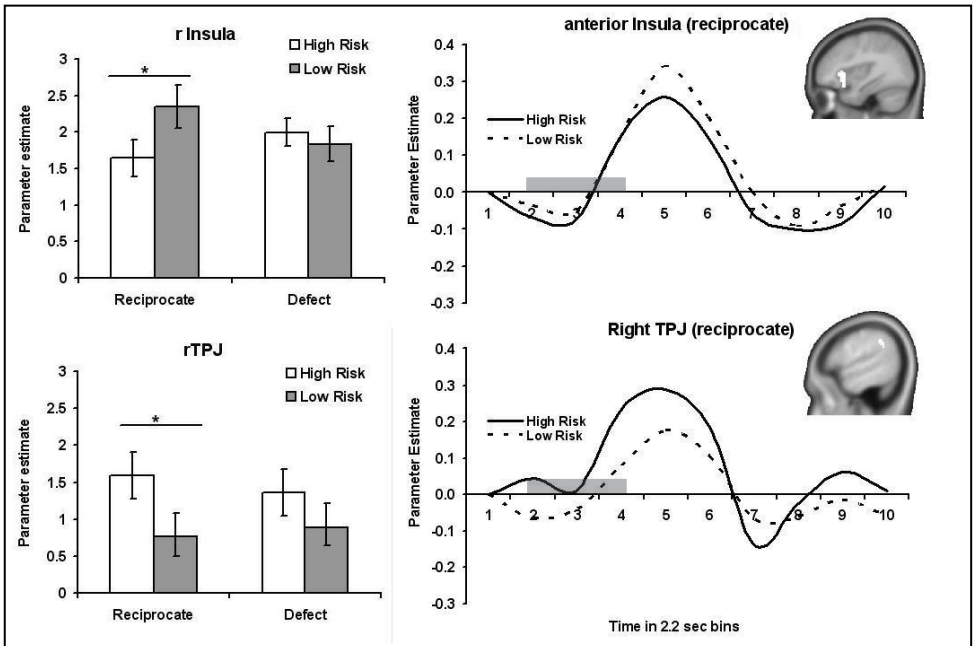


Figure 3.4: ROI parameter estimates and time series for regions that were sensitive to the risk manipulation (error bars represent standard error). rTPJ (MNI: 45, -43, 32) was more active for reciprocate choices when the risk that the first player took by trusting was high rather than low. In contrast, the right anterior Insula (MNI: 36, 24, 0) was more active for reciprocate choices when the risk that the first player took was low relative to high. The time-series plots show the data collapsed over conditions, gray area represent decision window of participant.

3.4 Discussion

The goal of this study was to investigate the neural correlates of reciprocity motives in brain regions that have previously been associated with mentalizing (aMPFC, rTPJ), reward and arousal (ventral striatum and insula) and inhibition of selfish impulses (ACC, rDLPFC). As expected, our behavioral results showed that participants reciprocated more when the first player took a high risk to trust and when the benefit of being trusted was high for the trustee, indicating that when reciprocating participants took into account both the consequences for the other as well as for themselves (Pillutla et al., 2003; van den Bos et al., manuscript submitted). Consistent with previous studies, our brain imaging data demonstrated that several brain regions worked together when individuals reciprocated trust and, in addition, provided more insight into how these regions were differentially sensitive to reciprocity motives.

First, separate analyses revealed that the two important areas of the mentalizing network, the aMPFC and rTPJ (Frith and Frith, 2003) have

separable functions in reciprocal behavior. Consistent with previous studies, the aMPFC was more active when participants defected compared to when they reciprocated (Gallagher et al., 2002; Decety et al., 2004). As such, the aMPFC was more active when the personal outcome of the decision was the greatest. This result is consistent with the hypothesis that the aMPFC is important for self-referential processing (Northoff et al., 2006; Ochsner, 2008) and with the interpretation that the aMPFC may havea general role in the evaluation or representation of reward information (Harris et al., 2007; van den Bos et al., 2007; Hampton et al., 2008). However, supplementary analyses revealed that the activation in aMPFC was not sensitive to the magnitude of personal gain (see supplementary data). Contrary to our predictions, there was no effect of the benefit manipulation on the activity in the aMPFC. Apparently, activation in the aMPFC is not directly sensitive to changes in cooperative intentions of the other player, but this region is sensitive to increases in personal outcome (defection). In future studies, it will be important to not only test motives for reciprocity, but also motives for defection.

In contrast to the aMPFC, the right TPJ was not sensitive to the type of choice but was sensitive to the risk manipulation when reciprocating. Activity in this area was higher when participants reciprocated when the risk was high rather than low. In the high-risk condition, the consequences of the participants' decision to reciprocate were fairly large for the first player compared to the low-risk condition. This finding indicates that, in line with our hypotheses, the rTPJ is involved in the shifting attention from the self to the other (Lamm et al., 2007) in order to distinguish between the consequences for self and other in a social decision-making paradigm (Lamm et al., 2007). This interpretation is consistent with a recently postulated hypothesis that argues that the rTPJ is involved in the reorientation of attention from self to other (Decety and Lamm, 2007; Mitchell, 2008).

Interestingly, our results also show that the activity in the rTPJ is sensitive to individual differences in SVO. That is, proself individuals showed more activation in the rTPJ when reciprocating, whereas prosocial individuals showed more activation in the rTPJ when defecting. Different processes may underlie these differences in neural activation for prosocials and proselfs, but one explanation may be that individuals with a prosocial orientation have their goals more aligned with those of the other, leading to less attention shifting when reciprocating, but more attention shifting when defecting (Decety and Hodges, 2006). These hypotheses should be further tested in future research.

The ventral striatum and insula were hypothesized to be sensitive to reward and arousal manipulations and were expected to be particularly sensitive to individual differences in reciprocal behavior. Indeed, regression analyses

demonstrated that activity in the striatum was higher for reciprocal choices than for defective choices for the prosocial participants (albeit at an unconservative threshold, but confirmed by unbiased ROI analyses, see supplementary results), whereas the proself participants showed the opposite pattern. The pattern of activation for the prosocial individuals is consistent with prior studies, which showed that cooperative choices are associated with ventral striatum activity (Fehr and Camerer, 2007). Even though the choice to reciprocate resulted in larger mutual gain, it also yieled a smaller monetary personal reward. Possibly, for prosocial individuals reciprocating in itself has a higher reward value whereas for proself individuals the personal gain has a higher reward value. This interpretation should be treated with caution, because it relies on reverse inferencing (Poldrack, 2006), but the results fit with a hypothesis postulated in a recent review analysis on other-regarding preferences (Fehr and Camerer, 2007). This hypothesis suggests that the ventral striatum represents the positive experienced utility of cooperation.

The insula was also sensitive to individual differences in SVO. However, the insula showed the opposite pattern of activity compared to the striatum. Furthermore, the insula showed sensitivity to the risk manipulation. The pattern of activation suggests that the insula is indeed sensitive to norm violations (King-Casas et al., 2008). That is, prosocial participants showed more activation in the insula when they defected (the unlikely alternative given their SVO), whereas the proself participants showed more activation in the insula when they reciprocated (again, the less likely option given their SVO). In addition, the insula was activated on those trials where participants chose to reciprocate when the risk that the first player took was low. In that case, there was less incentive to reciprocate than in the high risk situations. However, even though the choice to reciprocate occurred less frequently when the risk was low compared to when it was high, our supplementary analyses, using the frequency of the choice as covariate, revealed that these effects could not be attributed to a nonspecific effect of frequency. Together, these findings support the hypothesis that the insula is most active when a personal norm is violated (which can be a reciprocate norm for prosocial individuals or a defect norm for proself individuals) (Singer et al., 2006; Montague and Lohrenz, 2007). As such, the anterior insula have a more general role in social decision-making besides marking events as negative, such as pain, disgust or unfair offers (Sanfey et al., 2004; de Vignemont and Singer, 2006). Rather, the insula may be sensitive to the arousal associated with norm violations, which could also explain why the anterior insula are activated following other types of unexpected events such as a risk prediction error (Preuschoff et al., 2008). Alternatively, the insula

responses to violation of personal norms may serve as control signals, which mark social expectation violations (King-Casas et al.,2008).

Prior studies have suggested that cooperative behavior involves not only brain regions which are sensitive to mentalizing or reward representation, but also the control of impulses and actions. These studies have suggested that the ACC and the rDLPFC are important for regulating impulses to either defect or cooperate (Knoch et al., 2006; Rilling et al., 2007). Consistent with these earlier studies, in the current study, we showed that indeed the ACC and the rDLPFC were most active when social impulse control was required. In particular, ACC and rDLPFC were activated when participants reciprocated even though the benefit of being trusted was low. In other words, when the external incentive to reciprocate was low, the ACC and the rDLPFC were more engaged in reciprocal decisions. Inspection of the figures shows that the pattern of results observed for the insula follows a similar pattern as observed for ACC and rDLPFC, regions thought to be important for cognitive control (Ridderinkhof et al., 2004) and inhibition of self-oriented impulses (Knoch et al., 2006). It should be noted that, in this study, we could not distinguish between brain activity related to the actual choice and the appraisal of this choice. Thus, it is possible that ACC and rDLPFC activation is associated with the decision phase and the insula activation with the appraisal phase. These are important questions to test in future research.

Furthermore, activation in the ACC but not the rDLPFC, was also modulated by SVO. In prosocial individuals, the ACC was more active when reciprocating than when defecting, whereas in proself individuals, the ACC was more active when defecting than when reciprocating. One explanation for its role in both overriding the tendency to defect when the benefit is low, and the modulation of defecting vs. reciprocating depending on SVO, may be associated with the experience of response conflict (Botvinick et al., 1999). Importantly, activation in ACC and rDLPFC was not correlated with the frequency of making specific choices, arguing against the possibility that the effects can be explained by non-specific frequency effects.

Conclusion

Together, the results of this study demonstrated that several brain regions are differentially sensitive to reciprocity motives. We demonstrate that even though several brain areas are sensitive to individual differences in SVO (ACC, insula, rTPJ), these regions are differentially sensitive to the risk and benefit manipulations. The combined interpretation of sensitivity to SVO and modulation by risk and benefit manipulations allowed for advanced inference of the putative roles of these regions in reciprocal behavior. Our analyses revealed

the different motives for reciprocity, the risk for the trustor and the benefit for the trustee could be dissociated on the neural level. This study suggests a number of directions for future research as well as testable hypotheses. The differential involvement of the reported regions in reciprocal exchange demonstrates that neuroimaging methods may provide insight in the neural correlates of behavioral differences between individuals. It is possible that similar social interaction tasks could be used to explore social processing in a variety of populations, including developmental populations as well as individuals who fail to take the intentions of others into account.

3.5 Supplementary material

Supplementary Table 3.1: Percentage of average reciprocity in the four conditions. Standard errors between brackets.

	High Benefit	Low Benefit
High Risk	64% (.06)	53% (.06)
Low Risk	58% (.06)	27% (.04)

Supplementary Table 3.2: percentage of cooperative choices per factor level and participant. Participants in red are excluded from further fMRI analysis because there were either no observations in the reciprocate or defect condition.

PPN	High Risk High Benefit	High Risk Low Benefit	Low Risk High Benefit	Low Risk Low Benefit
101	9 (36%)	11 (48%)	17 (49%)	15 (36%)
102*	43 (100%)	44 (100%)	48 (100%)	5 (12%)
103	19 (76%)	17 (74%)	17 (49%)	21 (50%)
104	22 (88%)	19 (83%)	32 (91%)	26 (62%)
105	17 (68%)	11 (48%)	22 (63%)	18 (43%)
106	14 (56%)	8 (35%)	15 (43%)	23 (55%)
107	13 (52%)	7 (30%)	23 (66%)	6 (14%)
108	22 (88%)	20 (87%)	29 (83%)	12 (29%)
109*	0 (0%)	0 (0%)	0 (0%)	0 (0%)
110	8 (32%)	13 (57%)	4 (11%)	6 (14%)
111	23 (92%)	8 (35%)	29 (83%)	10 (24%)
112	12 (48%)	9 (39%)	9 (26%)	5 (12%)
113*	43 (100%)	44 (100%)	45 (94%)	25 (48%)
114	21 (84%)	17 (74%)	23 (66%)	27 (64%)
115	22 (88%)	12 (52%)	26 (74%)	10 (24%)
116	22 (88%)	19 (83%)	31 (89%)	11 (26%)
117	21 (84%)	16 (70%)	30 (86%)	9 (21%)
118	20 (80%)	8 (35%)	29 (83%)	6 (14%)
119	13 (52%)	10 (43%)	14 (40%)	5 (12%)
120*	0 (0%)	0 (0%)	0 (0%)	0 (0%)
121	7 (28%)	7 (30%)	6 (17%)	5 (12%)
122	5 (20%)	8 (35%)	12 (34%)	11 (26%)

Relation reciprocity level and activation in ROIs

The relation between individual differences in reciprocity and BOLD responses are further illustrated based on ROI values in Supplementary Figure 3.1. Based on the results of the regressions analysis the following reported results are all from the defect > reciprocate contrast. Subsequent post hoc analyses revealed that the reported effects were also significant for the reciprocate > fixation contrast but not for the defect > fixation contrast. Thus, these effects were primarily driven by differences in neural activation associated with reciprocity choices.

The patterns reported in Suppl. Figure 1 demonstrate a positive correlation between reciprocal behavior and activation in the anterior Insula (bilateral), the ACC and the right TPJ($r = .79$, $p<.001$, $r = .59$, $p<.005$ and $r = .65$, $p<.001$ respectively). These patterns demonstrate that the individuals who reciprocated less (pro-self oriented) recruited these areas more when reciprocating compared to defecting whereas the prosocial individuals recruited these areas more when defecting compared to reciprocating. .

In contrast, there was a negative correlation with reciprocity scores and activity in the ventral striatum ($r = -.64$, $p<.001$, see Supplementary Figure 3.1). These patterns show that the individuals who reciprocated more (prosocial individuals) activated these areas more when reciprocating,

Frequency effects

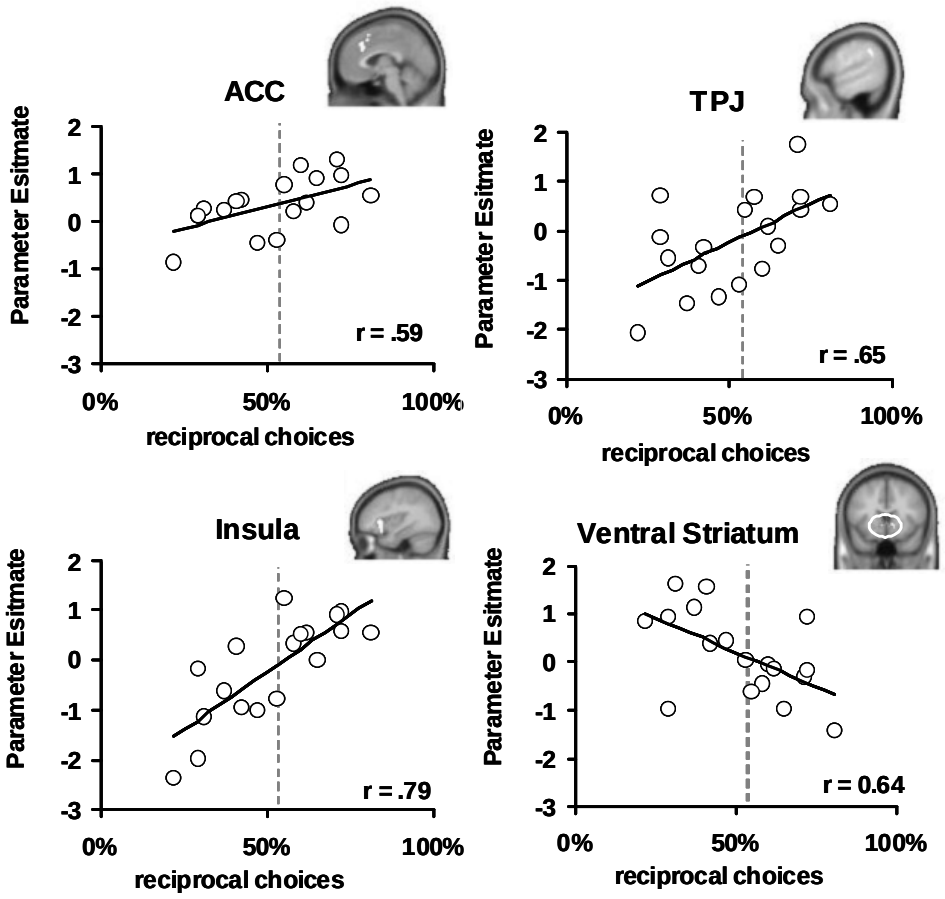
The results above show that the changes in activation can be influenced by frequency effects. To test whether the reported effects of Risk and Benefit correlated activation in the ROIs can be explained by frequency differences we performed several additional analyses.

First, we correlated the beta values for Low Benefit and High Benefit conditions with the frequency of these choices in the right DLPFC and the ACC. In all cases there was no significant correlation between the frequency of the choice and the beta values (all p's >.3). The same analysis for the anterior insula resulted in a negative correlation between High Benefit Reciprocate frequency and beta values for neural activation in this condition ($r = -.588$ $p < .05$) and for Low Benefit Reciprocate frequency and neural activation in this condition ($r = -.719$, $p < .001$). This negative correlation indicates that the less frequent a choice was made, the higher the beta value for that choice.

Finally, we analyzed the effect of High and Low Risk in the anterior insula and TPJ in a similar manner. There were no correlations between frequency and beta values in the TPJ and Insula (all p's >.1). This result was expected because activation in TPJ follows a pattern of activation opposite of frequency sensitivity.

Subsequently, we performed ANCOVAs with the frequency of the decisions as covariate. The results of these analyses show that all the effects reported in the original manuscript remain significant (at $p<.05$ threshold), furthermore the frequency covariates were never of significant influence (all p's >.1). There was again one exception and that was the anterior insula; when testing for the effects of risk and benefit with frequency as covariate we found that the frequency of high benefit trials was significant ($F(1,13) = 4.972$, $p< .044$).

Together, these effects show that risk and benefit effects are not correlated with the frequency of choices, except for neural activation related to benefit, but not risk, manipulations in the insula.



Supplementary Figure 3.1: Correlations between average level of reciprocity in brain activity in the defect > reciprocate contrast.

Reward Magnitude

To further explore our interpretation of aMPFC function we further analyzed its sensitivity to reward magnitude. The small difference in the pay-off between trials made it possible to look at the neural correlates of the relative difference in gain for defect choices. For this analysis we divided the defect trials into either low or high gain trials (the difference only being one coin, or ten cents). For these analyses we contrasted the high gain defect choices with low gain defected choices. These analyses did not yield any significant effects at our $p<.001$ & 12 voxels threshold. Lowering the threshold for exploratory reason did not resulted in activation in aMPFC.

4. Changing brains, changing perspectives: The neurocognitive development of reciprocity

Adolescence is characterized by the emergence of advanced forms of social perspective-taking and substantial changes in social behavior. Yet, little is known about how changes in social cognition are related to changes in brain function during adolescence. This study investigated the neural correlates of social behavior in three phases of adolescence using fMRI while participants played the second player in a Trust Game. With age, adolescents were increasingly sensitive to the perspective of the other player as indicated by their reciprocal behavior. These advanced forms of social perspective-taking were associated with increased involvement of the left temporal parietal junction (TPJ) and the right dorsolateral prefrontal cortex (DLPFC). In contrast, young adolescents showed more activity in the anterior medial prefrontal cortex (aMPFC), a region previously associated with self-oriented processing and mentalizing. These findings suggest that the asynchronous development of these neural systems may underlie the shift from self towards other-oriented thought.

4.1 Introduction

"*When I was a boy of 14, my father was so ignorant I could hardly stand to have the old man around. But when I got to be 21, I was astonished at how much the old man had learned in seven years."* (Arnett, 2000)

This quote by Mark Twain (1835-1910) illustrates the importance of understanding changes in perspective-taking across adolescence. Although this phenomenon has attracted attention for centuries, the question how these changes arise is still as debated today as it was 100 years ago. For example, it is well known that early in adolescence, individuals are still more inclined towards self-oriented thought and actions (Eisenberg, Carlo, Murphy, & Van Court 1995; Elkind, 1985), whereas later in adolescence individuals become more inclined towards thinking about others, taking social responsibility and controlling their impulses (Steinberg, 2009). Additionally, recent studies have

shown that functional changes occur in 'social brain' regions (for a review see Blakemore, 2008). It is, however, not yet known how changes in brain function contribute to specific changes in social behavior and perspective-taking. Understanding the emergence of social behavior and perspective-taking in adolescence is of high importance to society, as it is the critical transition period during which children gradually become independent individuals.

Recently, reciprocal exchange in social interaction has been examined with a simple economic exchange game; the Trust Game (Berg, Dickhaut, & McCabe, 1995) (see Figure 4.1). In the Trust Game two players can share a certain amount of money. The first player can choose to divide the money equally between herself and the second player, or to give it all to the second player with the advantage that the amount then increases in value. The second player has the choice to reciprocate and share the increased amount of money with the first player (act prosocial), or to defect and exploit the given trust by keeping most of the money for herself (act proself). This game touches on a central issue in the development of social perspective-taking; it requires the ability to understand intentions of and benefits for others.

Prior studies with adults using functional magnetic resonance imaging (fMRI) demonstrated different neural circuits for the receipt and the display of prosocial behavior in the Trust Game (King-Casas et al., 2005; Krueger et al., 2008; van den Bos, van Dijk, Westenberg, Rombouts, & Crone, 2009b). In particular, when the second player receives trust from the first player, a network of areas including the temporal parietal junction (TPJ) is activated. Several meta-analyses have shown that in social contexts the TPJ is important for shifting attention between own and other perspectives and inferring intentions (Mitchell, 2008; van Overwalle, 2009). It has therefore been suggested that within the context of the trust game, receiving trust might result in a shift in perspective from self to the other (King-Casas et al., Krueger et al., 2008, van den Bos et al., 2009b).

In contrast, a different network is activated when the second player decides to either reciprocate or exploit trust. In particular, anterior medial prefrontal cortex (aMPFC) activity has been reported when individuals exploit trust and maximize own gains (van den Bos et al., 2009b). This region has also been reported to be important for first players when they trust another individual, with the expectation of increasing their own pay-off (McCabe et al., 2001). It is suggested that the aMPFC activity in context of the Trust Game reflects the evaluation of own outcomes or thinking about one's reputation (Frith & Frith, 2008).

Thus, the TPJ and the aMPFC, which together have been described as part of the 'social brain' network (van Overwalle 2009), seem to have separable roles

in reciprocal behavior. Importantly, these regions work in concert with brain circuits which are important for regulation of thought and action such as the dorsolateral prefrontal cortex (DLPFC) (Miller & Cohen, 2001). In particular, the DLPFC was found to be important for the control of selfish or self-oriented impulses in several economic games (Knoch, Pascual-Leone, Meyer, Treyer, & Fehr, 2006; Rilling et al., 2007). Importantly, DLPFC is one of the brain regions that shows the most protracted structural as well functional development (Crone, 2009).

One of the predictions that follows from these prior studies is that adolescent development of perspective-taking in social decision-making is associated with different recruitment of aMPFC, TPJ and DLPFC. Our specific hypotheses about the neural developmental brain changes related to social behavior were informed by studies showing developmental changes in the brain during childhood and adolescence. In prior studies using simple tasks that involve thinking about different social scenarios, young adolescents showed less activity in TPJ, but increased activity in aMPFC compared to adults (Blakemore et al., 2007; Pfeifer, Lieberman, & Dapretto, 2007; Wang, Lee, Sigman, & Dapretto, 2006).We predicted that defecting (a self-oriented act) would be associated with increased aMPFC activity, given its role in thinking about self-motives relative to intentions and goals of others. Under the hypothesis that especially in early adolescence individuals are more inclined towards self-oriented thought and action (Eisenberg et al., 1995; Elkind, 1985), we predicted higher defection in early adolescents and more activity in self-related brain areas (aMPFC), relative to mid adolescents and adults. Furthermore, under the hypothesis that adolescents show late changes in intention consideration (Blakemore, 2008), we predicted that activity in TPJ when receiving trust would increase between early adolescence and adulthood. Finally, based on developmental studies that demonstrated increased activity in cognitive control and emotion regulation tasks with increasing age (Crone et al., 2006; Steinberg, 2005), we expected that DLPFC would be increasingly engaged during adolescence in intention consideration and reciprocity.

To test these hypotheses, we examined behavioral choices and neural responses of second players in the Trust Game in three age groups selected based on adolescent developmental stage; pubertal early adolescents (12-14 years), post-pubertal mid adolescents (15-17 years) and young adults (18-22 years). Based on our own and other behavioral studies with economic games, we expected an increase in the general level of reciprocity with age (Sutter & Kocher, 2007; van den Bos, van Dijk, Westenberg, & Crone, 2009a).

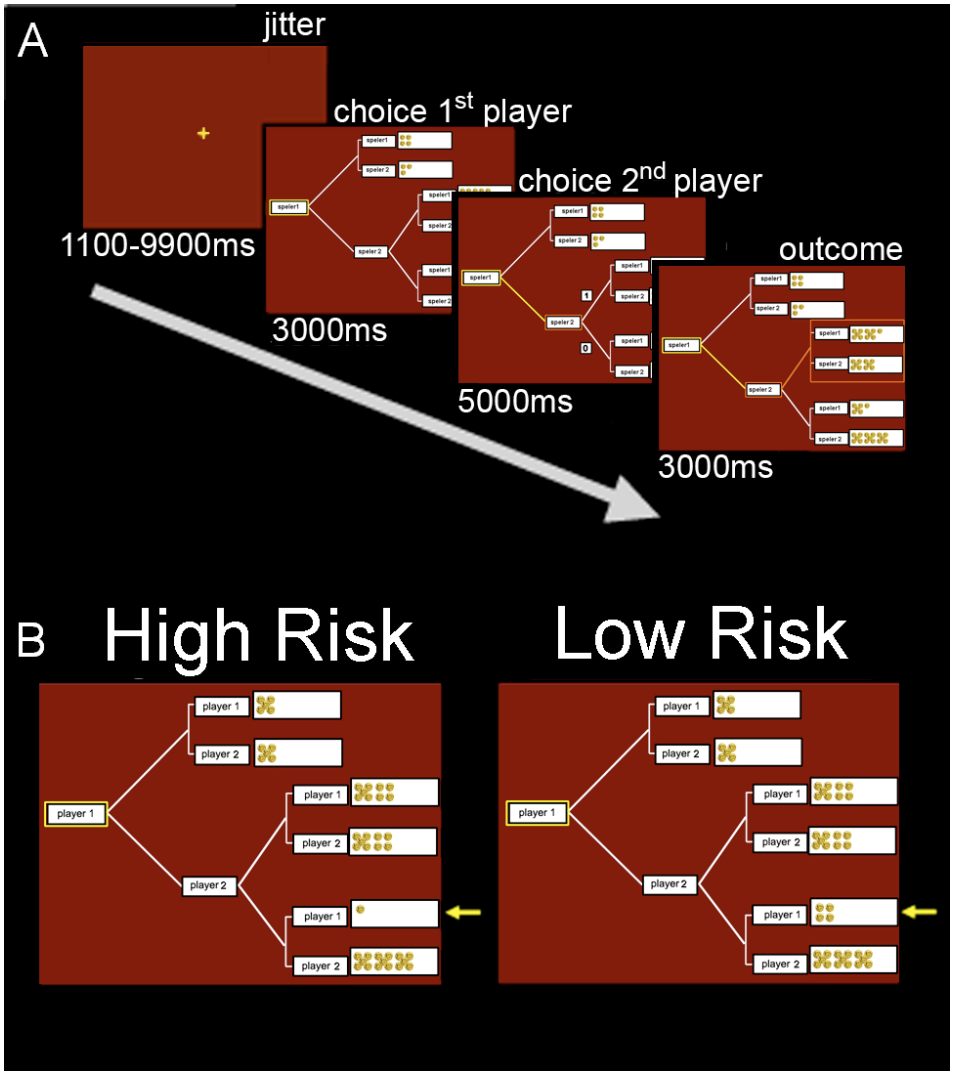


Figure 4.1: A: Each trial started with a 3-second display of the two choice alternatives for the first player; trust or no trust. After 3 seconds the trust or no-trust decision was shown to the participant. When the first player chose not to trust, the no-trust outcome was visually highlighted for 3- sec and the trial ended. For those trials on which the first player chose to trust, participants were instructed to make their decision within a 5-second window. The 5-sec decision-display was followed by either a 3-sec display of the outcome of their decision (reciprocate or defect) or a "too late" screen in case the participant did not respond within 5 seconds. In case of trust the total amount of money increased with a factor between 1.8 and 2.2.

To further test the ability to understand others' intentions, we added a task condition in which we manipulated the amount that the first player could lose by trusting the second player (the participant) (Malhotra, 2004; van den Bos et al., 2009a, 2009b, see Figure 4.1). In the analyses the trials on which the first player could lose a relatively large amount were labeled high-risk choices, and the trials on which the first player could lose only a small amount were labeled low-risk choices. Higher level of reciprocity in the high-risk context is hypothesized to reflect the recognition of the positive intentions of the first player, relative to the low-risk context (Malhtora, 2004; Pillutla, Malhotra, & Murnighan, 2003). As a consequence, this additional manipulation enabled us to obtain a behavioral measure of social perspective-taking within the task, with the expectation of larger risk-related reciprocity differentiation (RDS) for the older participants who are more capable of identifying intentions and integrating perspectives (van den Bos et al., 2009a).

4.2 Methods

4.2.1 Participants

Sixty-two healthy right-handed paid volunteers (30 female, 32 male; ages 12-22, M = 16.2, SD = 2.9) participated in the fMRI experiment. Eight participants were excluded from the fMRI analysis because they had an unreliable number of observations in one of the conditions ($n<4$). Age groups were based on adolescent development stage, resulting in groups composed of early adolescence/pubertal (12- to 14-year-olds, N=21, 11 females), mid adolescence/post-pubertal, (15- to 17-year-olds, N=15, 7 females) and young adults (18- to 22-year-olds, N=18, 9 females). A chi square analysis indicated that the gender distribution was similar across age groups ($X2\ (2) = .114$, $p = .94$). The data from the adults were also reported in another study (van den Bos et al., 2009b). Participants gave informed consent for the study, and all procedures were approved by the medical ethical committee of the Leiden University Medical Center (LUMC).

Participants completed the Raven Standard Progressive Matrices (R-SPM) for an estimate of their reasoning skills (Raven, 1941), and the Tanner scale (Tanner, 1975) for an estimate of their stage of pubertal development (see Table S4.1). There were no significant differences in IQ between the different age groups ($F\ (2, 51) = .62$, $p = .54$), and the Tanner stage development demonstrated a significant difference in puberty levels between age groups 12-14 (M = 2.95, SE = .24) and 15-17 (M = 4.11, SE = .22, $t(1,33) = 3.89$, $p < .001$).

4.2.2. Task Procedure

The procedure for the Trust Game was similar to the previously reported imaging study with adults (van den Bos et al., 2009b, see Figure 4.1). Participants were instructed that in an earlier phase of the study, other individuals had been assigned the roles of first player, and that they would complete the study in the role of second player inside the scanner. Furthermore, they were instructed that both the participant and the other players were financially rewarded based on the choices made during experiment. In each round of the experiment, participants were paired with a different, anonymous player who was matched for age and gender. At the end of the experiment the computer randomly selected the outcome of 5 trials and the sum of these trials determined the participants' payoff.

Unknown to the participant the decisions of the first player were not the decisions of real other participants, but were preprogrammed to reflect the behavioral pattern that was displayed in an earlier study (van den Bos et al, 2009a). In total, the task consisted of 145 trials; 96 trust trials and 49 no trust trials. The trials were divided over 4 blocks of 8.5 minutes each. The trials were presented in pseudo-random order with a jittered interstimulus interval (min=1.1-sec, max=9.9 sec, mean= 3.37 sec).

Before the experiment participants received a written explanation of the task, filled out a questionnaire and played 12 "practice" rounds. None of the participants failed this test.

4.2.3. fMRI Data Acquisition and Analysis

Data were acquired using a 3.0T Philips Achieva scanner at the LUMC. T2*-weighted EPIs (TR= 2.2 sec, TE= 30ms, 80 x 80 matrix, FOV = 220, 35 2.75mm transverse slices with 0.28mm gap) were obtained during 4 functional runs of 232 volumes each. A high-resolution T1-weighted anatomical scan was obtained from each participant after the functional runs. Data were analyzed using SPM2 (Wellcome Department of Cognitive Neurology, London). The functional time series were realigned, normalized to EPI templates, and spatially smoothed using a 8 mm full-width half-maximum Gaussian kernel. There were no significant differences in movement parameters between age groups ($F(2, 51) = 1.03$, $p = .36$).

Statistical analyses were performed on individual participants' data using the general linear model in SPM2. The fMRI time series data were modeled by a series of events convolved with a canonical haemodynamic response function (HRF). The start of the first player's choice display, no-trust and trust outcomes were modeled as 0-duration events. The trust outcomes were divided into reciprocate and defect decisions. These trial functions were used as covariates in

a general linear model, along with a basic set of cosine functions that high-pass filtered the data, and a covariate for run effects. The least-squares parameter estimates of height of the best-fitting HRF for each condition were used in pairwise contrasts. At the group level, contrasts between conditions were computed by performing one-tailed t-tests on these images, treating participants as a random effect. Results were considered significant at an uncorrected threshold $p > .001$ and $k > 10$ voxels.

We further performed voxelwise ANOVAs to identify regions that showed age related differences in relation to social decision-making. The developmental patterns in the behavior and fMRI data we constrained to a specific set of contrasts that captured developmental trends (linear increase [-1 1 0] ∩ [0 -1 1], early increase [-2 1 1], late increase [-1 -1 2], and their inverse) in the trust vs. no trust and defect vs. reciprocate comparisons. For the age analyses we used a more stringent threshold of $p < .0002$, using a Bonferroni correction for multiple comparisons ($p < .001 / 6$).

We used the MARSBAR toolbox for SPM2 (Brett et al., 2002) to extract BOLD activity time series in Regions of Interest (ROI) to further characterize patterns of activity. We created ROIs of the regions that were identified in the functional mask of whole brain analyses.

4.3 Results

4.3.1. Behavioral Results

Increasing effect of intentions on behavior. On average participants reciprocated about half of the trials (M = 53%), but there were large individual differences in behavior (SD = 17%, Min = 12%, Max = 87%; see Figure 4.2A). As predicted, the analyses of risk showed that participants reciprocated more when the risk for player 1 was high compared to when it was low (F (2, 51) = 25.22, $p < .001$, see Figure 4.2B). Even though there were no age related differences in mean reciprocal choices (F (2, 51) < 1, $p = .66$; see Figure 4.2A), there was an age x risk interaction for percentage of reciprocal choices (F (2, 51) = 5.44, $p < .007$, see Figure 4.2B). As expected, a post hoc Tukey test confirmed that all groups differed significantly from each other in RDS score at $p < .05$. Furthermore, only for the older adolescents and adults there was more reciprocity for high-risk than for low-risk trials (both p's < .01), whereas the youngest adolescent group did not differentiate between high- and low-risk trials ($p = .8$, Figure 4.2B).

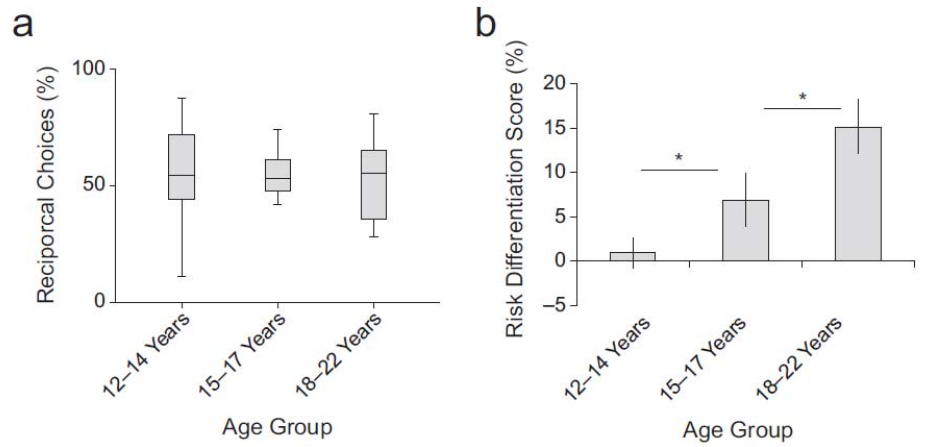


Figure 4.2: A) Average level of reciprocity for the total duration of the experiment displayed separately for each age group. Error bars represent standard deviation. Testing for outliers at 95% confidence interval did not yield any outliers for any of the age groups. **B)** The risk differentiation score (high-risk reciprocity – low-risk reciprocity) for each age group. This graph shows that the RDS increased with age, a post hoc Tukey test revealed that all groups differed significantly from each other ($p < .05$). Error bars are standard errors.

4.3.2 fMRI Results

*Receiving Trust***.** To identify the neural correlates of receiving trust, which was hypothesized to be associated with consideration of the intentions of the other, we compared the [Trust – No Trust] contrast across all participants. This analysis revealed increased activity in a large network of areas associated with cognitive control; the DLPFC, parietal cortex and dorsal medial frontal cortex/anterior cingulate cortex (ACC) (see Table 1). Subsequently, we tested the hypothesis of age related changes in activity related to receiving trust by performing mixed linear and non-linear ANOVAs with age group as between participant factor. As anticipated, the conjunction contrast [-1 1 0] ∩ [0 -1 1] demonstrated age related changes in left TPJ. Additionally, the contrast [-1 -1 2] revealed activity in right DLPFC (see Figure 4.3, Table 4.1). Time-series analyses of l-TPJ showed heightened activity for both reciprocate and defect choices compared to no-trust trials, however this difference was not significant in early adolescence, whereas it was present for late adolescents and greatest for the young adults (see Figure 4.3). In contrast, the time series analysis for DLFPC revealed heightened activity for reciprocate and defect choices relative to no-trust trials only for the young adults. The correlations between individual risk difference scores (RDS) and activity in these areas ($r = .37$, $p < .006$ for l-

TPJ and r = .45, p < .001 for r-DLPFC, see Figure S4.1) strengthens the hypothesis of a relation between l-TPJ and r-DLPFC function and intention identification and perspective-taking.

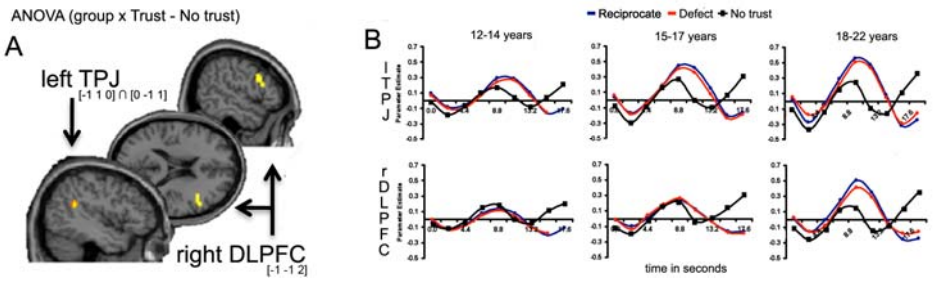


Figure 4.3: Age differences in l-TPJ and r-LPFC activity associated with receiving trust. **A)** Activation maps with the clusters that show significant linear [-1 1 0] ∩ [0 -1 1] and non-linear [-1 -1 2] increases in trust related activation with age. **B)** Showing the time-course of activation in the l-TPJ and the r-DLPFC related to defect, reciprocate and no trust. Along the x-axis 0 seconds indicates 1) the onset of the choice for the participant in case of trust, or 2) the outcome of the experiment in case of no trust.

*Defect vs. Reciprocate***.** Next, we investigated the neural correlates of proself versus prosocial motivated acts, by examining differences in neural activity for reciprocate and defect choices following trust outcomes. As expected, the [Defect – Reciprocate] contrast across all participants revealed increased BOLD response in the aMPFC (Figure 4.4 and Table 4.1). Additional activity was found in the left anterior Insula and the right inferior frontal gyrus. Consistent with our previous findings (van den Bos et al., 2009b), the opposite contrast [Reciprocate – Defect] did not result in significant changes in neural activity.

To further investigate whether there were age related changes in [Defect - Reciprocate] activity, we performed linear and non-linear ANOVAs with age group as between subjects factor on the [Defect – Reciprocate] contrast. The contrast [-2 1 1] revealed an age related change which was specific for the aMPFC (see Figure 4.4 and Table 4.1). These findings demonstrate that the differential engagement of the aMPFC increases between early and mid adolescence and then remains stable in mid to late adolescence/early adulthood.

The time-series of the aMPFC region revealed increased activity compared to baseline for defect choices in all age groups. Closer inspection of the activation patterns revealed that early adolescents also demonstrate heightened activity for reciprocal choices compared to baseline. Thus, consistent with the hypothesis of heightened aMPFC activity in early adolescence, we demonstrate a *decrease* in aMPFC activity related to reciprocal choices with age. This was

further confirmed by a significant negative age correlation for reciprocate > fixation (r = .56, p < .02). No such correlation was observed for defect > fixation (r = .06, p = .72).

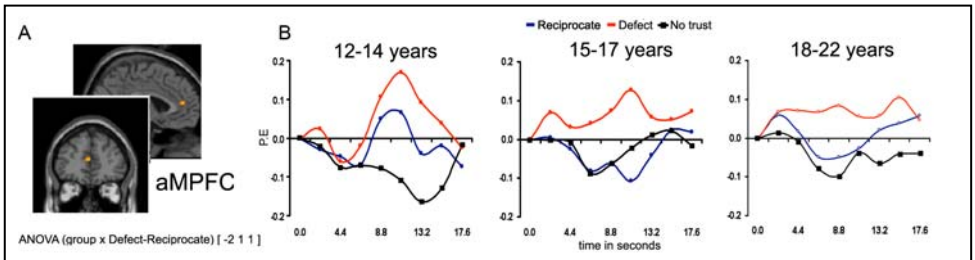


Figure 4.4: Age differences in aMPFC activity associated with defect vs. reciprocate decisions. **A)** Activation maps with the clusters that show an early increase [-2 1 1] in the difference in defect vs reciprocate activation. **B)** Showing the time-course of activation in the l-TPJ and the r-DLPFC related to defect, reciprocate and no trust. Along the x-axis 0 seconds indicates 1) the onset of the choice for the participant in case of trust, or 2) the outcome of the experiment in case of no trust.

Table 4.1: Brain Regions revealed by whole brain contrasts**.**

Anatomical region	L/R	vxls	Z	MNI coordinates		
				x	y	z
Receiving Trust						
[Trust - No Trust]						
Superior Parietal Lobule	R	71	4.14	21	-66	54
Precuneus	L	121	4.18	-30	-45	42
Caudate / Dorsal Striatum	L/R	431	5.20	-15	0	15
ANOVA [Trust - No Trust] [-1 1 0] ∩ [0 -1 1]						
TPJ	L	44	4.06	-44	-46	29
ANOVA [Trust - No Trust] [-1 -1 2]						
DLPFC	R	56	4.01	44	16	21
Choice Type						
[Defect – Reciprocate]						
anterior Medial Prefrontal Cortex	L/R	774	4.89	0	42	6
Visual Cortex	L/R	733	8.82	6	-93	12
Insular Cortex	L	63	4.82	-36	24	-12
Inferior Frontal Gyrus	R	27	3.95	62	21	0
[Reciprocate – Defect]						
Visual Cortex	L/R	490	7.72	6	-73	6
ANOVA [Defect – Reciprocate] [-2 1 1]						
anterior Medial Prefrontal Cortex	L/R	78	5.84	2	42	15

MNI coordinators for main effects, peak voxels reported at p < .001, at least 12 contiguous voxels. Age contrasts were corrected for multiple comparisons; p < .001 / 6. For each ROI, the center of mass is reported.

Individual differences. A final question concerned the relation between neural activity and the average level of prosocial behavior displayed in the task. A whole-brain regression analyses on the [Defect – Reciprocate] contrast with average reciprocity per individual as predictor revealed activation in bilateral anterior Insula, dorsal anterior cingulate cortex (dACC) and r-DLPFC (Table S4.2, Figure 4.5). Higher reciprocity was thus associated with more activation in these areas when defecting, and higher defection was associated with more activation in these areas when reciprocating.

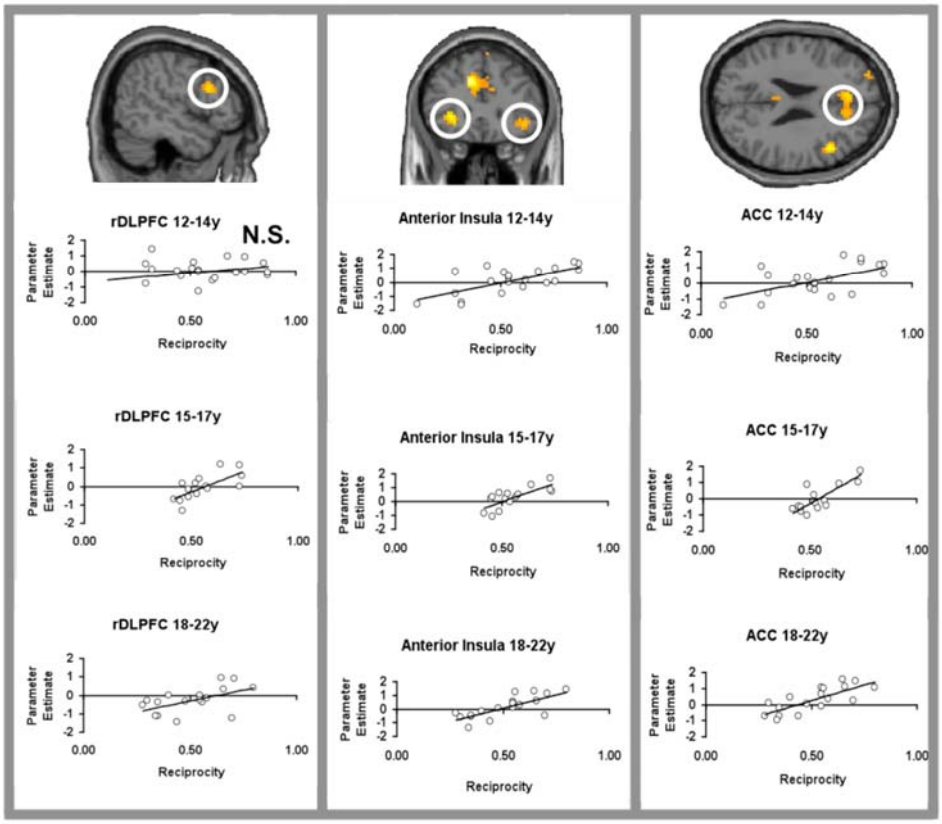


Figure 4.5: Activation maps for the regression analysis on the [Defect – Reciprocate] contrast with average level of reciprocity as covariate for all participants, threshold at $p < .001$. Separate scatter plots representing the correlations between the [Defect-Reciprocate] parameter estimate and average reciprocity for each age group separately, all based on the ROIs extracted from the whole group regression analysis.

4.4 Discussion

We investigated adolescence as a transitional period, during which linear as well as non-linear changes in social reasoning and associated brain circuitry take place (Casey et al., 2008). Indeed, analyses of age differences demonstrate that the regions implicated in social behavior followed asynchronous developmental patterns, with faster maturation of aMPFC but late maturation of l-TPJ and r-DLPFC. This asynchronous pattern of functional brain development may bias adolescents towards different social behavior in daily life (Casey et al., 2008; Paus, Keshavan, & Giedd, 2008; Steinberg, 2005).

The behavioral data are consistent with prior observational studies which marked adolescence as a transition period for social behavior (Eisenberg et al., 1995, 2005). Interestingly, these results highlight that adolescence is not necessarily characterized by general increases of prosocial behavior, but rather by an increase in the sensitivity to the perspective of others in social decision-making (see also Blakemore 2008; Kohlberg, 1981; Selman 1980). That is, increased consideration of consequences for others (i.e., increased RDS) was accompanied by both an *increase* in reciprocity on high-risk trials and a *decrease* of reciprocity for low-risk trials, and importantly the youngest adolescents did not show sensitivity to the perspective of the other. Alternatively, the age related increase in risk differentiation could be the result of increased inequity aversion (Fehr & Schmidt 1999). Both explanations are consistent with the notion of advanced forms perspective-taking in adolescence.

Our reasoning that receiving trust was associated with more active deliberation of the motives of others was further supported by increased activity in the l-TPJ, an area that is implicated in taking the perspective of others and inferring intentions (Mitchell, 2008; van Overwalle, 2009). In support of the hypothesized shift in attention from self to the other during adolescence, we observed an increase in the engagement of the l-TPJ with age. Moreover, the suggested role of the l-TPJ in shifting perspective from self to other was further supported by the correlation between l-TPJ activity and the behavioral index of perspective-taking (RDS); the more participants differentiated between the low an high-risk context, the more active the l-TPJ was after receiving trust. In addition, the pattern of activation of the l-TPJ, and the absence of an effect of risk on behavior for the youngest adolescents, suggests that in early adolescence focus of attention is not (yet) on the outcomes and intentions of others, and that there are still changes between mid adolescence and young adulthood in the focus on the other. These findings are in line with prior social scenario reading studies, which also demonstrated an increase in the l-TPJ activity between ages 10-18 and 22-32-years (Blakemore et al., 2007). Furthermore, recent studies revealed that TPJ is correlated with self reports of altruism (Tankersley et al.,

2007) and charitable giving (Hare et al., 2010), consistent with the presumed role of shifting attention from self to others in a social context.

Besides activity in the l-TPJ, we found that young adults, when receiving trust, showed increased activity in the r-DLPFC, an area previously found to be involved in tasks requiring cognitive control (Miller & Cohen, 2001) and the control of selfish or self-oriented impulses in context of social dilemmas (Rilling et al., 2007). This activity may indicate a regulatory role of r-DLPFC in social exchange as it was more active for adults for the non-preferred response alternative (Knoch et al., 2006). Consistent with studies which employed cognitive control paradigms (Crone et al., 2006) our results indicated an increase in the engagement of the r-DLPFC with age. Apparently, over the course of adolescence not only the development of the l-TPJ, but also the r-DLPFC contributes to a refinement in social behavior, which is supported by the finding that activity in the r-DLPFC also correlated with the ability to infer intentions of others (risk difference score). Thus, the differential involvement of l-TPJ and r-DLPFC marks mid adolescence (15-17-years) as an important transition period for intention consideration and social behavior, during which not all children are yet recruiting the associated brain regions to the same extent as adults, but during which emerging intention consideration is on its way.

If the changes in social behavior are associated with increased consideration of the outcomes for the other, what then motivated adolescents to act selfish? What are the neural correlates of self-oriented behavior? These questions were tackled by the comparison of defect and reciprocate choices which revealed increased activity in the aMPFC for defect choices in young adults and mid adolescents. Given the role of the aMPFC in processing self-referential and self-relevant events (for a review see van Overwalle, 2009), these findings suggest that participants were more involved in self-oriented thought when they defect and thus maximize personal outcome. The question then arises; how does this region support self-oriented acts in early adolescence; do adolescents show increased activity for defect choices? Intriguingly, this was not the case. When acting pro-self (i.e., when defecting), early adolescents showed similar activity in aMPFC as mid adolescents and young adults. When reciprocating, however, young adolescents also showed activity in aMPFC. This activity was not found in mid adolescents and adults. One of the fascinating questions for future research is to test the hypothesis that even when reciprocating young adolescents are engaged in self-referential thoughts. Prior research has demonstrated that in late childhood/early adolescence, social interaction is considered from an egocentric perspective (Eisenberg et al., 1995; Elkind, 1985). Possibly it is not until mid adolescence that a prosocial act becomes more automatic and less self-engaged.

Although meta-analyses of social cognition for adults (Lieberman, 2007; van Overwalle 2009) and adolescents (Blakemore, 2008) have indicated the importance of the aMPFC in self referential processes, other research has implicated this region in mentalizing, or thinking about what others are thinking about you (Amodio & Frith, 2006). In particular, in the context of social interactions the role of the aMPFC has been related to considering one's reputation (Frith & Frith, 2008). Future studies should unravel which of these aspects of self-referential processing is changing in early to mid adolescence.

This study brings us a step closer towards understanding why Mark Twain started to understand his father better when he was 21 than when he was 14. Most likely this was associated with increased perspective-taking skills subserved by interacting brain regions important for social reasoning. Future research could benefit from analyzing connectivity between these areas to better understand how these regions contribute to social behavior (Burnett & Blakemore, 2009). Finally, prior studies have shown that the combined use of neuroimaging and game theoretical paradigms can further the understanding of the neural underpinnings of psychopathology (Chiu et al., 2009).Therefore, the current findings on normative social development can also be the basis for understanding the development of psychopathology in adolescence (Paus et al., 2008).

4.5 Supplementary Material

Table S4.1. Group scores for IQ, reaction times (RT), head movement and gender distribution (SD = Standard Deviation; mm= millimeter). None of the group differences are significant (all *p*'s > .5)

Group differences			
	12-14 years	15-17years	18-22years
Raven IQ (SD)	121.2 (5.2)	121.5(6.2)	119.2(8.1)
RT in seconds (SD)	1.6(0.6)	1.7(0.6)	1.8(0.5)
Movement (mm)	0.75	0.76	0.73
Female (Male)	11(10)	7(8)	9(9)

Table S4.2. : Brain Regions revealed by regression analysis

Anatomical region	L/R	voxels	Z	MNI coordinates		
				x	y	z
Regression [Defect – Reciprocity] w/ avg. reciprocity			Z			
anterior Cingulate Cortex	L/R	335	4.70	-9	27	36
anterior Insula	R	241	5.12	33	21	0
	L	133	4.72	-33	24	0
Superior parietal cortex	R	150	4.04	21	-66	54
DLPFC	R	84	4.38	48	18	24

MNI coordinates for main effects across all participants, peak voxels reported at $p < .001$, at least 12 contiguous voxels.

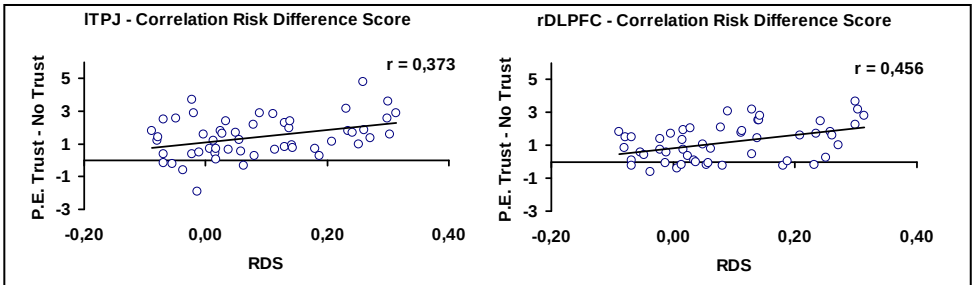


Figure S4.1: Plots showing the correlation between parameter estimates for the Trust - No Trust contrast and the risk differentiation scores (RDS), in the l-TPJ and the r-DLPFC.

5. Dissociable brain networks involved in development of fairness considerations

In this functional magnetic resonance imaging study, we examined developmental changes in the brain regions involved in reactions to unfair allocations. Previous studies on adults suggested that reactions to unfairness are not only affected by the distribution itself but also by the ascribed intentionality of the proposer. In the current study, we employed the mini Ultimatum Game (Falk, Fehr, & Fischbacher, 2003) to examine responder behavior to unfair offers of varying degrees of intentionality. Sixty-eight participants from four age groups (10-, 13-, 15-, and 20-year-olds) carried out the task while fMRI data were acquired. Replicating previous findings in adults, participants of all ages showed activation in the bilateral insula and dorsal anterior cingulate cortex (dACC) during rejection of unintentional but acceptance of intentional unfair offers. Rejection of unintentional unfair offers involved increasing activation with age in the temporoparietal junction and the dorsolateral prefrontal cortex. These findings provide evidence for an early developing insula-dACC network involved in detecting personal norm-violations and gradually increasing involvement of temporal and prefrontal brain regions related to intentionality considerations in social reasoning. The results are discussed in light of recent findings on the development of the adolescent social brain network.

5.1 Introduction

Fairness consideration is a key component of social interactions and involves the comparison between outcomes for self and other. People prefer equitable distribution of resources and react strongly to inequitable distributions, which has also been termed as inequity aversion (Fehr & Schmidt, 1999). In this sense, fairness forms a socially shared norm. Violations of norms, behaviors that deviate from the norm, are generally perceived to be aversive, where people want to be nice to those who treat them fairly and hurt others who do not treat them fairly (Fehr & Schmidt, 1999). However, assessment of behaviors that deviate from the norm goes paired with a second process assessing its intentionality (Falk et al., 2008; Fehr & Schmidt, 1999). For example, Blount

(1995) showed that behavioral reactions to unfairness are strongly modulated by the ascription of intentionality: people react less negative to disadvantageous inequity when they feel the inequity was not intentional. This process of intentionality understanding requires the ability to mentalize about other individuals' goals and intentions. In human development, behavioral studies have suggested that inequitable distribution of resources (i.e., unfairness) is aversive from an age as early as 7-8 years (Fehr et al., 2008), followed by increased understanding of intentionality in adolescence (Güroğlu et al., 2009; Selman, 1980; van den Bos et al., 2010). The goal of this study was to examine the development of the neural correlates of intentionality understanding related to fairness considerations.

Neuroscientific studies have identified separable brain regions involved in these different aspects of fairness considerations. These studies typically employ the Ultimatum Game (Güth et al., 1982), where two players are given a stake to share. The first player (the proposer) makes an offer that the second player (the responder) can accept or reject. Acceptance of the offer results in sharing the stake between the two players as proposed, whereas rejection of the offer yields both players to go empty-handed. On the one hand, functional magnetic resonance studies using the Ultimatum Game suggest that bilateral insula activation might reflect the detection of norm violations following unfair proposals (Güroğlu et al., 2010; Sanfey et al., 2003). In addition, transcranial magnetic stimulation and neuroimaging studies suggest that the dorsolateral prefrontal cortex (DLPFC) might be important for overriding self-interest (accepting unfair offers in an Ultimatum Game) and thereby enable participants to act upon their inequity aversion, or violation of the fairness norm (Knoch et al., 2010; Knoch et al., 2006a; van 't Wout et al., 2005).

On the other hand, considering others' intentions involves the activation of the temporoparietal junction (TPJ) (Frith & Frith, 2007; van Overwalle, 2009). Activity in this region has been related to switching attention between different perspectives (Mitchell, 2008) and is also involved in competitive games (Assaf et al., 2009; Halko et al., 2009; Polezzi et al., 2008) and charitable giving (Hare et al., 2010). A neuroimaging study with adults showed that the insula, DLPFC and TPJ had dissociable patterns of activation during a fairness game which allowed for the separation of processes involved in fairness considerations (Güroğlu et al., 2010). In sum, neuroimaging findings suggest that the insula might be involved in detecting social norm violations, the DLPFC in the regulation social behavior (e.g., rejection of unfair offers), and the TPJ in intentionality considerations.

Brain regions such as TPJ and DLPFC show protracted structural development (Gogtay et al., 2004), suggesting that the ability to understand

intentions and the control of selfish impulses mature relatively late. Indeed, recent behavioral and neuroimaging studies provide support for the development of perspective taking (Dumontheil et al., 2009) and the contribution of the TPJ to social reasoning across adolescence (Sebastian et al., 2008; van den Bos et al., 2011). In previous behavioral research, we demonstrated that the ability to judge fairness develops at an early age, whereas the ability to understand intentions does not develop fully until late adolescence (Güroğlu et al., 2009).

Accordingly, we hypothesized that that the slow emergence of intentionality consideration in fairness judgments is associated with protracted development of the DLPFC and TPJ. Using the mini-Ultimatum game, we examined intentionality understanding in unintended versus intended unfair offers. We predicted that responses to unintended unfair offers would require increased intentionality consideration and regulation of social behavior, and therefore would be associated with increased DLPFC and TPJ activation that emerges gradually over adolescence. Further, we hypothesize that TPJ activity might be increased during the rejection of unintentional offers, because the participants might then make additional considerations about what the proposer might think about their rejection, which is generally not considered to be the socially acceptable decision (Güroğlu et al., 2009).

5.2 Methods

5.2.1 Participants

Sixty-eight participants from four age groups took part in the study: 10-year-olds (N = 17, M age = 10.4, SD = 0.86; 6 females), 13-year-olds (N= 15, M age = 13.4, SD = 0.51; 8 females), 15-year-olds (N= 13, M age = 15.4, SD = 0.51; 5 females), and 20-year-olds (N= 23, M age = 20.4, SD = 1.67; 13 females). Gender distribution was similar across age groups ($\chi 2(3) = 2.39$, $p = .50$). The data from the young adults have been previously reported (Güroğlu et al., 2010). All participants were healthy and right-handed volunteers without neurological or psychiatric impairments. All participants provided informed consent; participants younger than 18 years-old were accompanied by their parents who also provided consent. A radiologist reviewed all anatomical scans; no anomalies were found.

In order to obtain an estimate of intelligence, 10-year-olds completed two subscales (Block design and Similarities) of the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1991), 13- and 15-year-olds completed the same subscales of the (revised) adult version, the Wechsler Adult Intelligence Scale (WAIS-R; Wechsler, 1997) and 20-year-olds completed the Raven Standard

Progressive Matrices (Carpenter et al., 1990). The scores were converted to Intelligence Quotient (IQ) estimates and participants had average IQ ($M = 107.93$, $SD = 11.53$); there were no significant age differences ($F(3, 66) = 1.69$, $p = .18$) and IQ scores did not correlate with behavioral performance in terms of rejection rates of unfair offers (all $r(67) < 0.14$, $p > 0.27$).

5.2.2 Task description

Participants played the role of the responder in the modified version of the Ultimatum Game (UG) which incorporates intentionality considerations (Güroğlu et al., 2009; Güroğlu et al., 2010). In this version, the first player (proposer) is presented with a fixed set of two distributions for sharing the stake (here 10 coins) with the responder (i.e., the second player). There were three conditions in the game; in each condition one of the distributions was an unfair distribution of the stake with 8 coins for the proposer and 2 coins for the responder (i.e., 8/2 offer). The three conditions were termed depending on the alternative offer pitted against the 8/2 offer: a) 5/5 offer (fair-alternative), b) 2/8 offer (hyperfair-alternative), and c) 8/2 offer (no-alternative).

Participants practiced the task (24 trials) on a computer before the scanning session and subsequently they played 168 trials of the game with anonymous age and gender matched partners. These 168 trials consisted of 126 trials of unfair offers (42 per condition, 3 conditions: fair-, hyperfair-, and no-alternative) and 42 alternative offers (21 for fair- and hyperfair-alternative conditions each). The trials were presented in three blocks of 42 trials lasting about 8.3 min each.

Each trial started with the presentation of the fixation cross followed by the presentation of the set of offers available to the proposer, where the offer made by the proposer was encircled in red, and the Yes and No buttons (see Figure 5.1). Participants could accept or reject the offer by pressing a button using the index and middle fingers of their right hand. If they failed to respond within 5000 ms, a screen displaying ‘Too late!’ was presented for 1000 ms. Upon responding, the response was presented on the screen until the end of the 6000 ms. Trials were randomized and presented with a jittered interstimulus interval (mean = 1530 s, min = 550 ms, max = 4950 ms; optimized with OptSeq2, surfer.nmr.mgh.harvard.edu/optseq/, developed by (Dale, 1999)).

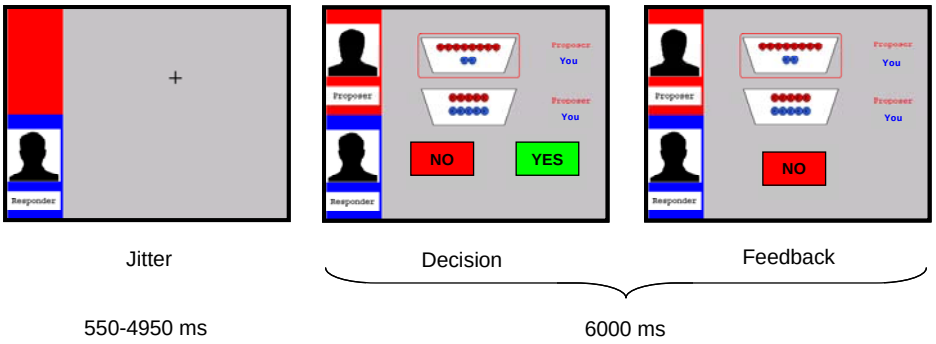


Figure 5.1: Visual display of events presented in the scanner task. Trials started with a jittered fixation screen lasting 550-4950 ms. The left panel in the decision screen displayed the name of the proposer in red (here 'proposer') and the name of the responder (here 'responder'). Two offers each containing red and blue coins indicate the share for the proposer and the responder, respectively (here 8/2 vs 5/5) and the offer made by the proposer was encircled in red (here 5/5). The responder was a maximum response time of given 5000ms to select Yes or No to accept or reject the offer. Upon response, the feedback screen displayed the given response (here 'Yes') until 6000 ms after the start of the trial.

Each trial was played with a new player to avoid learning and reputation effects. Only the first name and the first letter of the surname of the players were displayed on screen to ensure anonymity. Participants were told that the offers of the proposers had already been obtained in a previous part of the study and that at the end of the session the computer would randomly select ten trials that would determine their total earnings. In order to emphasize the interactive character of the game with consequences for them and the other players, participants were explained that the proposers' earnings would be contingent upon their decisions. At the end of the session, a screen was presented indicating the pay-off (five euros for each participant). In reality, the offers presented to the participants were computer simulated but were based on behavior reported in prior experiments (Güroğlu et al., 2009). After the scan session, none of the participants expressed doubts about the cover story.

5.2.3 MRI data acquisition

The scanning session was carried out at the university medical center using a 3.0T Philips Achieva. Using E-Prime software, stimuli were projected onto a screen at the head of the scanner bore and participants viewed the stimuli by means of a mirror mounted on the head coil assembly. The scanning sessions

consisted of four types of scans in the following order: i) localizer scan, ii) T2*-weighted echo-planar imaging (EPI) sequence measuring the bold-oxygen-level-dependent (BOLD) signal (TR= 2.2 sec, TE= 30ms, slice-matrix= 80 x 80, slice-thickness=2.75mm, slice gap = 0.28mm gap, field of view (FOV) = 220 mm), iii) high-resolution T1-weighted anatomical scan, and iv) high resolution T2-weighted matched-bandwidth high-resolution anatomical scan with the same slice prescription as the EPIs. Each of the three blocks of functional runs consisted of 200 volumes; the first two scans were discarded to allow for equilibration of T1 saturation effects.

5.2.4 MRI data analysis

SPM5 software (www.fil.ion.ucl.ac.uk) was used for image preprocessing and analyses. Slice-time correction, realignment, spatial normalization to EPI templates, and spatial smoothing using a 8mm full-width half-maximum 3D Gaussian kernel were carried out. The youngest age group moved significantly more than the other three age groups (main effect of Age $F(3, 67) = 3.21$, $p < .05$, followed by posthoc Tukey comparisons). However, the total amount of movement was minimal: the maximum movement parameters were below 1.81 mm for all participants and all scans. The functional time series were modeled by a series of events convolved with a canonical haemodynamic response function (HRF). The moment of stimulus presentation with zero duration was used to model the data. For the purposes of this study, the unfair offers (8/2 offers) were modeled separately based on context (3 levels: fair-, hyperfair-, or no-alternative) and response (2 levels: accept or reject). Contrast images for each individual were used in the second-level random effects model to run full-factorial analysis of variance and one-tailed post hoc t-tests. We further conducted regression analyses to test for brain-behavior relations using mean rejection levels per condition. Unless otherwise indicated, the fMRI analyses were conducted at the commonly used (Sanfey et al., 2003; Tabibnia et al., 2008) threshold of $p < .001$ uncorrected with a voxel threshold of 10 functional voxels. Results are reported in the MNI305 stereotaxic space.

5.2.5 Region-of-interest (ROI) analyses

In order to further examine the effects obtained in the whole-brain full factorial ANOVAs, Region of Interest (ROI) analyses were conducted using the MARSBAR tool in SPM5 (Brett et al., 2002). These analyses were conducted in predetermined brain regions of interest, including the insula, the DLPFC and the TPJ.

5.3 Results

5.3.1 Behavioral results

A repeated-measures ANOVA was conducted with context (3 levels: fair-, hyperfair-, and no-alternative) as the within subjects factor, age (4 levels: 10-, 13-, 15-, and 20-year-olds) as between subjects factor and rejection rates of unfair offers as the dependent variable. There was a main effect of context ($F(2, 128) = 67.67$, $p < .001$) as well as a context x age interaction ($F(6, 128) = 3.00$, $p < .01$) (see Figure 5.2). Rejection rates of unfair offers in the fair-alternative condition were highest, followed by the hyperfair-alternative ($M = .79$, $SD = .25$ and $M = .73$, $SD = .27$, respectively; $F(1, 67) = 3.04$, $p = .05$), and lowest rejection rates were observed in the no-alternative condition ($M = .35$, $SD = .36$; $F(1, 67) = 73.58$, $p < .001$).

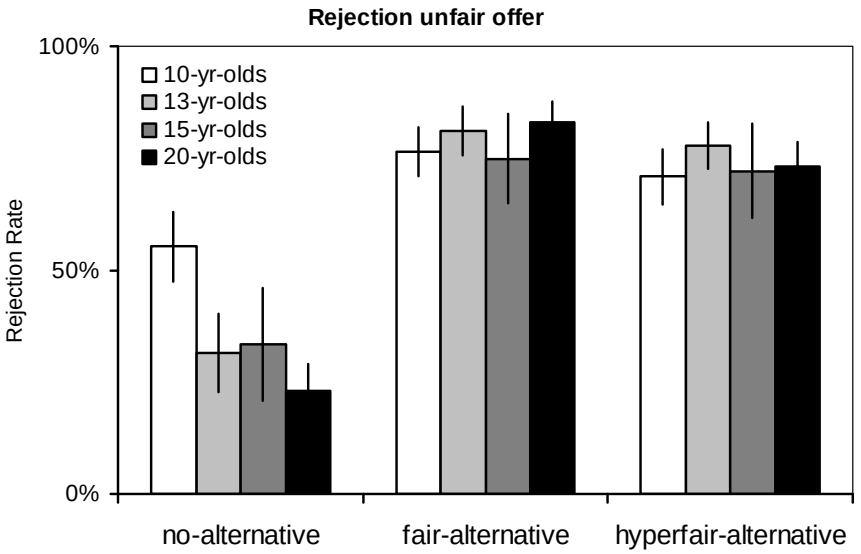


Figure 5.2: Display of means and standard deviations of rejection rates of unfair offers in the three conditions for the four age groups.

Tukey post-hoc analyses exploring the age x context interaction showed that rejection rates of unfair offers did not differ across age groups in the fair- and hyperfair-alternative conditions (both $F(3, 64) < .37$, $p > .78$) whereas they did in the no-alternative condition ($F(3, 64) = 2.90$, $p < .05$). Youngest participants rejected unfair offers in the no-alternative condition more often than oldest participants did ($M = .55$, $SD = .32$ and $M = .23$, $SD = .27$, respectively). Thirteen and 15-year-olds rated in between and did not differ from either age group ($M = .32$, $SD = .34$ and $M = .33$, $SD = .45$, respectively).

5.3.2 fMRI results

Response x Intentionality Interaction across ages. First, we examined developmental differences in the role of intentionality (i.e., context) in responses to unfairness[9]. Whole brain analyses conducted with a 2 x 3 x 4 full factorial ANOVA with response (2 levels: accept / reject) and context (3 levels: fair- / hyperfair- / no-alternative) as the within subject factors and age (4 levels: 10-, 13-, 15-, and 20-year-olds) as the between subject factor yielded no three-way interaction between response, context and age. There was a response x intentionality interaction across all age groups ($F(2,350) = 7.34$, FDR $p < .05$, 10 voxel threshold) in the dorsal ACC (MNI -3, 27, 36) and bilateral insula/inferior frontal gyrus (IFG; MNI (42, 24, -6 and -36, 15, -9), see Figure 5.3A). To further examine the interaction effect, ROI analyses were conducted in the three regions involved in the interaction. These post hoc analyses showed that the activation in both the bilateral insula/IFG and dorsal ACC were higher during rejection than acceptance of unfair offers in the no-alternative condition (all F (1, 48) > 8.95, $p < .004$), but higher during acceptance than rejection of unfair offers in the fair- and hyperfair-alternative conditions (all F (1, 49) > 7.79, $p < .007$ and F (1, 52) > 8.86, $p < .004$, respectively). These effects were found for all age groups, suggesting that these areas are sensitive to the response x intentionality interaction independent of age (see Figure 5.3B). In previous studies these brain regions are shown to play a role in personal norm violations, that is, related to behaviors that are not frequently displayed by the individual (Güroğlu et al., 2010; van den Bos et al., 2009). The role of these areas in personal norm violations was further supported by brain-behavior correlations. BOLD activity for the reject > accept contrast correlated negatively with mean rejection levels of unfair offers in the no-alternative (left insula $r = -.35$, $p < .05$), fair-alternative (right insula $r = -.32$, $p < .05$) and hyperfair-alternative condition (left insula $r = -.46$, $p = .001$, right insula $r = -.39$, $p < .01$, and dACC $r = -.44$, $p = .001$). In other words, participants who often accepted unfair offers (i.e., had low rejection rates) showed high levels of insula and/or dACC activity when they rejected these offers and vice versa.

[1] Main effects of response and intentionality were also explored. Examination of the main effect of response yielded significant activation in bilateral Insula (MNI -33, 18, -15 and 51, 15, 6; p< .001, 10 voxel threshold) for the Acceptance > Rejection contrast. There were no regions involved in the Rejection > Acceptance contrast (see Supplementary Table 1). Examining the main effect of intentionality, we only found activation in the occipital lobe (MNI 21, -96, 6; p< .001, 10 voxel threshold) for the fair-alternative > no-alternative condition. See supplementary table for main effect of offer type (unfair > fair offers) per intentionality condition.

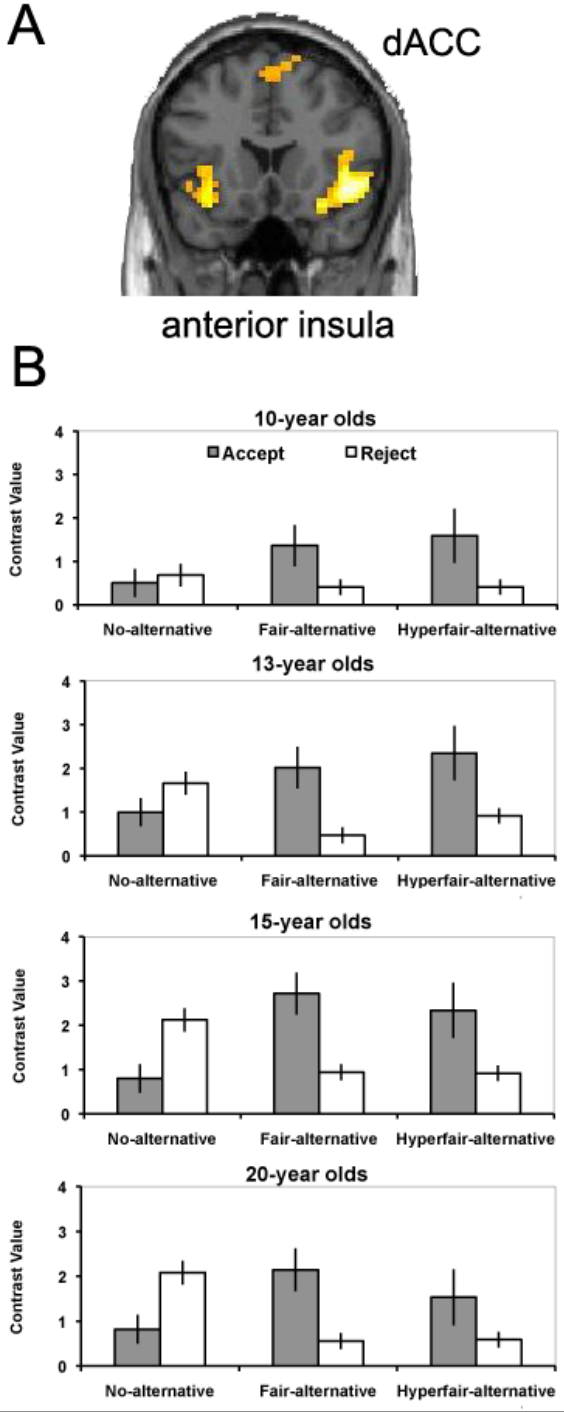


Figure 5.3: A) Results of the whole brain 2 (response) x 3 (context) interaction, showing the dACC [MNI -3, 27, 36] and bilateral insula/IFG [MNI 42,23, -6] at $p < .001$ 10 voxel threshold. **B)** Contrast values in the right insula for acceptance and rejection of unfair offers in the three conditions for the four age groups. Results for left Insula and dACC showed similar patterns of activity but are not shown.

Age differences in rejection in the no-alternative condition.

In order to examine developmental patterns in unintended versus intended unfair proposals we focused our analyses on brain areas that were specifically involved in rejection of unfair offers in the no-alternative condition with age included as a regressor in two separate contrasts.

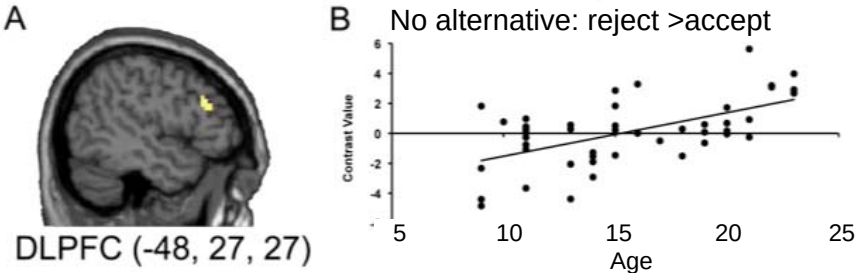


Figure 5. 4: A) Activation in the left dorsolateral prefrontal cortex (DLPFC; MNI -48, 27, 27) with positive correlation with age in the rejection no-alternative > acceptance no-alternative contrast; $p < .001$, 10 voxel threshold. **B)** Plot of contrast values for age and activity in left DLPFC for the rejection no-alternative > acceptance no-alternative contrast.

For the rejection > acceptance contrast in the no-alternative condition, brain activity in the DLPFC (MNI -48, 27, 27) correlated positively with age ($r = .57$; $T(60) = 3.23$; see Figure 5.4A and 5.4B). Other areas of activation are listed in Table 5.1. There were no negative correlations with age and no brain areas were correlated with age for the rejection versus acceptance contrasts in the hyperfair- and fair-alternative conditions. Thus, the age related increase in the DLPFC response was specific for no-alternative rejection relative to no-alternative acceptance trials. When no-alternative rejection behavior was added as covariate to the contrast, the DLPFC effect remained, showing that the effects are specific to age and cannot be solely explained on the basis of behavioral differences.

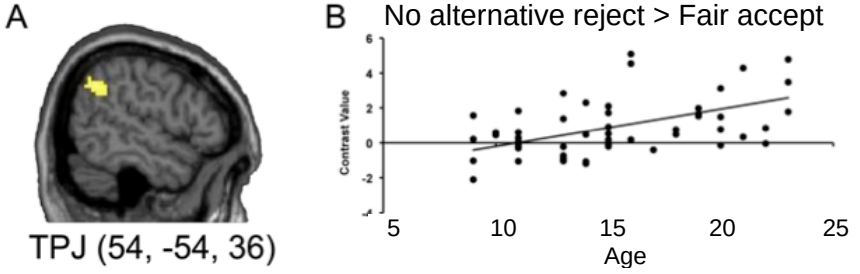


Figure 5.5: A) Activation in the right temporoparietal junction (TPJ; MNI 54, -54, 36) with positive correlation with age in the rejection no-alternative > rejection fair-alternative contrast; $p < .001$, 10 voxel threshold. B) Plot of contrast values for age and activity in right TPJ for the rejection no-alternative > rejection fair-alternative contrast.

Second, age was added as a regressor in the rejection no-alternative > rejection fair-alternative and rejection no-alternative > rejection hyperfair-alternative contrasts. Both contrasts resulted in positive correlations between BOLD activity and age in the TPJ (MNI 54, -54, 36 and 57, -48, 33, respectively; r = .51 and r = .50, respectively; T (53) = 3.25; see Figure 5.5A, 5.5B and Table 5.1). Other areas of activation are listed in Table 1. There were no negative correlations with age. Thus, age related increase in TPJ response was again specific for the no-alternative rejections relative to other types of rejections.

Mediation Analyses

To further investigate the relation between age, rejection rates in the no-alternative condition, and brain activity in DLPFC and TPJ we have performed mediation analyses. According to Baron and Kenny (1986), mediation can be established by demonstrating that (a) there is a direct effect of the independent variable (i.e., age) on the dependent variable (i.e., punishment), (b) there is a significant effect of the independent variable on the proposed mediator (i.e., anger), (c) the proposed mediator is correlated with the dependent variable after controlling for the independent variable, and (d) the effect of the independent variable on the dependent variable drops significantly when the mediator is included in a simultaneous regression (Baron & Kenny, 1986). First we investigated the mediation effect of DPLFC activity. As can be seen in Figure 5.5A, almost all the Baron and Kenny requirements are met. First, there is a significant effect of age on rejection rate (β = -.02), $t(49) = -2.01$, $p < .05$, and on the proposed mediator, contrast value [DLPFC reject – accept] (β = .29), $t(49) = 4.8$, $p < .001$. Second, DLPFC activity was borderline significantly correlated with rejection rate when controlling for age (β = -.05), $t(49) = -2.0$, $p = .05$. Third, the direct effect of age on rejection rate was no longer significant (β =-.006), $t(49) = -.54$, $p = .6$, when controlling for DLPFC activity. Finally, a Sobel test indicated that this reduction in significance was marginally significant, suggesting at least partial mediation (Sobel $z = -1.81$, $p = .07$).

Next we investigated the mediation effect of TPJ activity. As can be seen in Figure 5.5B, all the Baron and Kenny requirements are met again. First, there is a significant effect of age on rejection rate (β = -.03), $t(55) = -2.59$, $p < .02$, and on the proposed mediator, contrast value [TPJ reject_no-alternative – reject fair-alternative] (β = .21), $t(55) = 4.3$, $p < .001$. Second, TPJ activity was significantly correlated with rejection rate when controlling for age (β = -.07), $t(55) = -2.3$, $p < .03$. Third, the direct effect of age on rejection rate was no longer significant (β =-.1), $t(55) = -1.13$, $p = .26$, when controlling for TPJ

activity. Finally, a Sobel test indicated that this reduction was significant, suggesting full mediation (Sobel z = -1.9, p < .05).

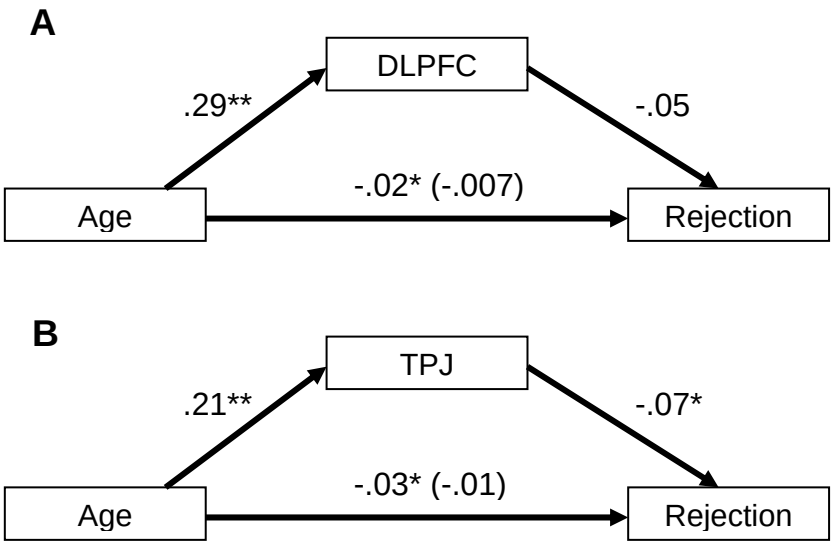


Figure 5.5 (**A**) Beta coefficients in the model testing for the mediation effect of neural activation in DLPFC for the rejection no-alternative > acceptance no-alternative contrast for the link between age and rejection of unfair offers in the no-alternative condition. (**B**) Beta coefficients in the model testing for the mediation effect of neural activation in TPJ for the rejection no-alternative > rejection fair-alternative contrast for the link between age and rejection of unfair offers in the no-alternative condition. ** p < .01, * p < .05.

5.4 Discussion

The goal of this study was to gain a better understanding of the emergence of intentionality understanding in fairness considerations. Using the mini Ultimatum Game we were able to distinguish between responses to unfair offers of varying degrees of intentionality. Consistent with prior behavioral studies, participants rejected unfair proposals when the alternative for the proposer was a fair division (Güth et al., 1982). This behavior has previously been reported across age groups and shows that fairness perceptions already play an important role in social decisions in late childhood and early adolescence (Fehr et al., 2008; Güroğlu et al., 2009; Sutter, 2007). However, the gradual emergence of intention-consideration in late childhood and adolescence was demonstrated by a decrease in rejection rates for unintentional unfair offers over the course of adolescence, with lowest rejection rates in adulthood. The results of this study

thus provide further support for improving intentionality understanding across adolescence (Güroğlu et al., 2010).

Importantly, we demonstrated that two different brain networks involved in fairness considerations develop at different rates and contribute to behavior in separate ways. First, a norm-violation network, including the anterior insula and the dorsal ACC, which develops relatively early in childhood, and second, a social brain network, including DLPFC and TPJ, which develops gradually over the course of adolescence, play a role in social decision-making involving fairness considerations. The developmental patterns of these networks set the stage for the interpretation of brain maturation during fairness considerations.

Early maturation of the norm violation network

Consistent with prior studies, anterior insula and dorsal ACC were differentially sensitive to acceptance and rejection responses, depending on the norm regarding the participant's behavior in the particular context, as defined by intentionality (Güroğlu et al., 2010). Namely, the activation of this network was related to acceptance of intentional unfair offers (i.e., in the context of a fair alternative where normative behavior would be to reject), but also to rejection of unintentional unfair offers (i.e., in the context of no alternative where normative behavior would be to accept). It should be noted here that the *norm violation* here is not to be confused with the detection of a social norm violation, which would be responses to unfair offers in general. Our findings show that perception of an unfair offer and the performed 'normative behavior' is highly context dependent. In this sense, the way we refer to norm violations is closer to *personal norms*, which are self-based standards of behavior in specific situations and differ from general attitudes or social norms referring to internalized self-expectations (Schwartz, 1977; Schwartz & Fleishman, 1978). This interpretation is strengthened by the correlations between brain activation and individual task behavior. That is, the dACC and insula network response when rejecting an unfair offer where the proposer had no alternative was even stronger for individuals who mostly accepted these offers. This role of the insula in personal norm violations is also supported by the relation between insula activity during social norm violations and individual differences in Machiavellianism (Spitzer et al., 2007) and social value orientation (van den Bos et al., 2009). Furthermore, the general function of this network in detecting deviations from the personal norm is supported by several studies showing its involvement in betrayals of trust (van den Bos et al., 2009) as well as in non-social norm violations such as risk prediction errors (Montague & Lohrenz, 2007; Singer et al., 2009). In this sense, the neural network including the anterior insula and dorsal ACC is related to behavior that deviates from personal

standards that are shaped by what one normally does within a particular context, that is, accepting an unfair offer in the no-alternative context and rejecting an unfair offer in the fair- and hyperfair-alternative contexts.

One limitation of the current study, and of social decision-making studies in general, is the relative low number of trials involved in the analysis. We should note that the analyses involving the acceptance of unfair offers in the fair- and hyperfair-alternative conditions may be suffering from low power, particularly in adults. The average number of trials for these conditions was relatively low (8.67 and 11.19, respectively). Although we have replicated our findings in an analysis which controlled for the number of trials, this is an issue that needs to be addressed in future research.

Notably, the norm-violation effects in the insula and dorsal ACC were observed for all age groups, showing that norm-violation are already detected by this network in young children. Indeed, behavioral studies have reported that already at age 7-8-years there is a strong preference for social norms of strict equity (Fehr et al., 2008) and a basic understanding of fairness (Güroğlu et al., 2009). It has been known for a long time that the rules for appropriate behavior are learned at a young age, as is shown by children's concepts of social rules (Piaget, 1956). The current findings indicate that children also rely on the insula / ACC network when judging their own social behavior in a particular context. These findings further suggest that the brain network related to fairness considerations including contextual information mature relatively early. However, the late maturing social brain network seems to incorporate extra information regarding intentionality into the decision-making process.

Late development of the social brain network

A crucial aspect of fairness considerations relates to our judgments of others' intentionality. Prior work has demonstrated that understanding intentions is associated with activation in the TPJ (Assaf et al., 2009; Halko et al., 2009; Polezzi et al., 2008; van Overwalle, 2009). These regions have also been implicated in inference of mental states (Hampton et al., 2008) and redirection of our focus of attention to others (Mitchell, 2008). In the current study, we hypothesized that TPJ was specifically associated with the considerations of unfair offers when the proposer did not have an alternative. Whereas children and adolescent showed similar activation of the insula and dorsal ACC as adults when rejecting no-alternative offers, TPJ involvement emerged gradually across adolescence. The intentions of the proposer are least clear in the no-alternative condition, which makes it likely that this condition exerts the highest mentalizing and intention consideration demands. Furthermore, the increased involvement of TPJ was specific for rejection of unfair offers in the no-

alternative condition. Whereas rejection of an unfair offer in the fair-alternative condition can be readily justified, this is not the case in the no-alternative condition. The consideration of self-interest and the related desire to reject an unfair offer, combined with the simultaneous (and automatic) consideration for lack of intentionality of the offer in this condition might also lead to feelings of guilt. Possibly, TPJ activation is related to these feelings of guilt towards others (Takahashi et al., 2004). This hypothesis needs further testing in future research.

In a pioneering set of studies, Blakemore and colleagues (Blakemore, 2008; Dumontheil et al., 2009; Sebastian et al., 2008) showed that the TPJ is less active in adolescents than adults during tasks requiring mentalizing. The current findings are consistent with these previous studies, and show that TPJ involvement is context-dependent. Furthermore, older adolescents are increasingly better able to take context, and thus intentionality-related information, into account while making decisions.

Besides TPJ, DLPFC was also more active during rejection of unintentional unfair offers in adults than in children, with an intermediate pattern for adolescents. In prior research, the slow maturation of DLPFC has been related to the emerging ability to control thoughts and actions (Bunge & Wright, 2007; Crone, 2009). Considering that the social norm is to accept unfair offers when there was no alternative, the increased DLPFC activation for rejection may indicate that adults override the tendency to accept (Knoch et al., 2006b). The negative correlation in children may indicate the opposite tendency; children may be inclined to reject unfair proposals (regardless of intentionality) and acceptance of unfair offers may require increased control. This interpretation should be tested in future research.

Finally, mediation analyses importantly demonstrated the mediating role of neural activity in the link between age and rejection rates of unfair offers. As such, these findings contribute to an understanding of the developmental mechanisms underlying age related changes in behavior. Our results suggest that age related differences in neural activation are partially responsible for behavioral differences that vary with age. Future longitudinal studies that incorporate structural brain development in the social brain network are crucial for further understanding of the mechanisms underlying development.

A new direction in understanding the development of fairness considerations

Two advantages of the current approach in examining development of social decision-making relative to prior reports is that we 1) included participants of four age groups, which is uncommon in fMRI studies, but allows for more precise measurement of developmental change (Galvan, 2010), and 2) related changes social brain network activation to real social behavior. Prior studies on

the development of the social brain network have typically involved comparisons of two groups (adolescents versus adults) whereas our approach allowed us to assess gradual changes over time. In addition, relative to prior studies, the current approach reveals that it is important to relate thinking about fairness and moral scenario's to actual social behavior in context, as behavior in the current task was modulated by intentionality considerations.

In sum, the current approach demonstrated development of the dissociable brain networks contributing to social decision-making across childhood, adolescence and adulthood. Regions associated with norm-violations showed a different developmental trajectory in their involvement in social decision-making than regions associated with perspective taking and intentionality consideration. The latter finding strengthens the claim that detection of norm-violations related to inequity and intentionality considerations are dissociable components of fairness consideration.

Finally, in future studies it is important to distinguish between different interaction partners in social interactions. In prior fMRI work in adults, it was demonstrated that interactions with friends was related to differential activation of a set of regions, including the ventral medial prefrontal cortex, the striatum and the amygdala (Güroğlu et al., 2008), and these regions may work together with the norm-detection and social brain networks reported here (e.g., Hare et al., 2010). Considering age differences in the social brain network (Blakemore, 2008), it is important in future research to understand how quality of relationships modulate the development of brain activation in social interactions across adolescence.

6. Who do you trust? Age comparisons of learning who to trust or distrust in repeated social interactions

Abstract

How do people learn to trust or distrust others? In a repeated trust game setting, we investigated the development of trust within repeated interactions. In addition to this relation-specific development of trust, we also assessed the development of trust across different age groups, ranging from late childhood to young adulthood. The results demonstrated that within relations, people use a tit-for-tat like strategy, but this pattern was more pronounced at a young age. With increasing age both the anger towards and punishment of non-cooperative players decreased. Further analyses showed that the differential willingness to punish violations of trust was mediated by feelings of anger.

6.1 Introduction

Trust plays an important role in almost all types of social interaction. In our daily lives trust is important in our relationships with family and friends, but also in many economic transactions with anonymous others (Rotter, 1967). Trust is important at all levels of society; it is often considered the 'glue' that holds society together (Fukuyama, 1995; Sullivan & Transue, 1999; Zak & Knack, 2001). Furthermore, trust is recognized to be of great importance for the development of social functioning throughout life, promoting moral behavior (Wright & Kirmani, 1997) and academic achievement (Imber, 1973; Wentzel, 1991).

Whereas the findings above indicate that trust is important for both social development and society, the origins of trust are still largely unknown. How do we learn to trust or distrust persons and anonymous institutions? To study the development of trust, social psychology has made extensive use of the Trust Game (TG). In the TG there are two players who can share a certain amount of money. The first player (trustor) has the possibility to divide a sum of money equally or to give it all to the second player (trustee). If the first player decides

to share the money, both players get their equal share and the game ends. However, if the first player gives all the money to the second player the total amount of money is tripled. Next, the second player has the possibility to reciprocate trust and share the increased amount of money with first player, or to exploit trust by keeping all the money (see Figure 6.1). It is clear that in the TG, player 1 faces the challenging question of whether or not to trust player 2: Will he/she reciprocate an act of trust? In more general terms, such an act of trust can be defined as the "willingness to make oneself vulnerable to others' actions based on a certain expectation of positive reciprocity" (Colman, 2003).

Insights from the research in which people played multiple Trust Games has revealed that trust in others generally increases after positive trust experiences, and decreases after experienced violations of trust (Delgado, Frank & Phelps, 2005; King-Casas, Tomlin, Anen, Camerer, Quartz & Montague, 2005; King-Casas, Sharp, Lomax-Bream, Lohrenz, Fonagy & Montague, 2008; De Cremer, Van Dijk & Pillutla, 2010). Whereas these findings address an important aspect of the development of trust, they do not address the general trust people may have in others. For example, these findings do not inform us to what extent people trust others in a first encounter, when they have not yet received any feedback on trustworthiness of the specific interaction partner. Interestingly, research shows that the initial trust in others is often high (Berg, Dickhaut & McCabe, 1995; Dufwenberg & Gneezy, 2000; McCabe, Houser, Ryan, Smith, & Trouard, 2001). Thus, in one-trial settings people often show high levels of trust, and in repeated settings they often show trust on the first encounter.

These findings not only show that it is important to distinguish between general trust and relation-specific trust, but also raise the question of how people's decisions on the first encounter are best explained. Here, social psychology has tended to focus on individual differences (e.g., differences in generalized trust, Yamagishi, Cook, & Watabe, 1998), and acknowledged that general trust may be shaped by people's personal histories of social interactions in the past (Rotter, 1967). In addition to these social psychological insights, it is interesting to see that recent research within developmental psychology, using one-trial game paradigms (Sutter & Kocher, 2006; Harbaugh, Krause, Liday & Vesterlund, 2002; van den Bos, Westenberg, van Dijk & Crone, 2009), has indicated that adults – i.e., the typical participants in social psychology studies - have higher levels of general trust compared to children. However, although these studies provided useful insights in how general trust changes with age, they do not inform us on age related changes in 'relation-specific' trust.

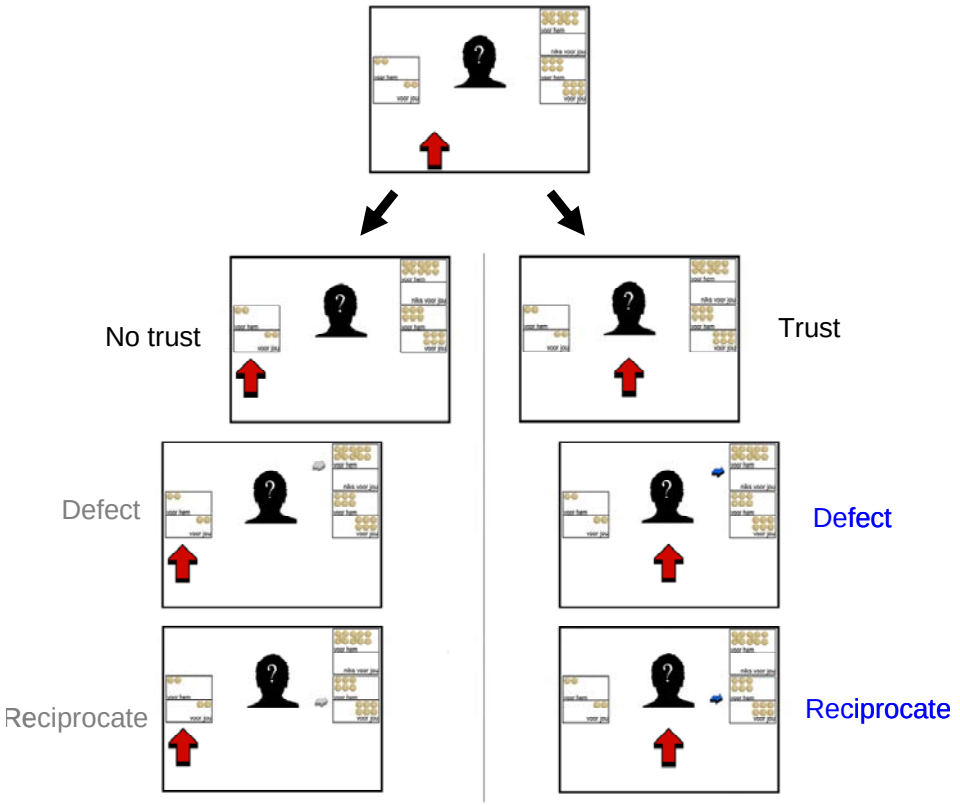


Figure 6.1: The sequence of visual displays that represent the different stages of the Repeated Trust Game (DTG). Each round starts with the identification of the other player. Next, the participant can decide to either trust or not to trust the other player. Regardless of the decision of the participant the choice by the other player to reciprocate or defect will be revealed. If the player decided to trust the choice by the other player is indicated with blue arrows and represent the real outcome of the game, whereas if the participant decided not to trust these arrows are displayed in grey and represent the counterfactual outcome of the game. Note that in the real experiment photographs of other players were shown, in stead of the silhouette with a question mark

Learning to trust and distrust

So how does trust develop? From the above, it is clear that to answer this question; we should distinguish between the general trust and relation-specific trust. It is also clear that social psychology and developmental psychology have each addressed different parts, but currently, the literatures more or less stand alone. To provide a more comprehensive picture of the development of trust, we therefore set out to integrate insights from both fields. For this purpose, and to investigate how people across different ages learn who to trust or distrust we use a repeated Trust Game paradigm in which participants from different ages

(children, adolescents, and adults) interact with the same players for several rounds (King-Casas et al., 2005). Because, as Rotenberg (1980) emphasized, it is equally important to learn who *not* to trust as to learn who to trust, the participants in the current experiment interacted with three different preprogrammed personalities that displayed different levels of trustworthiness (low, medium and high). During the repeated interactions the participants were playing the role of the trustor, thus each round they had to decide whether or not to trust the other. Following the number of trust decisions of the participants, we were able to study how the level of trust for each player changed based on the outcome of a series of social interactions.

In our studies, we distinguish between differences in general trust, which is observed at the first encounter in which one does not have any specific information about the interaction partner, and the subsequent 'relation-specific' changes in trust over time that occur after one has received feedback about the decisions made by one's interaction partner. Based on previous developmental studies with one-shot games we expect that with age participants will show higher levels of general trust and thus will be more prone to start the interaction with a trust move (Berg et al., 1995; Sutter & Kocher et al., 2008; van den Bos et al., 2009).

Subsequently, based on the outcomes of the social interaction with the three different players we expect that participants will learn how trustworthy each of the players is, and will act accordingly. Studies with similar paradigms have shown that adults often play a forgiving tit-for-tat like strategy (Wedekind & Milinski, 1996, Nowak & Sigmund, 1992). That is, if the other player shared in the last round the participant will trust in the following round (positive reciprocity), and if the other player did not share in the previous round the participant will react by not trusting in the next round (negative reciprocity). However, the forgiving tit-for-tat like strategy deviates from strict tit-for-tat by showing less negative reciprocity. Thus, adults may decide to trust the other even when that person did not share in the previous round, based on a history of positive reciprocity (Milinski & Wedekind, 1998).

We propose that although children display low levels of general trust on the first encounter, they are able to learn to trust *and* distrust their interaction partners based on a series of interactions. However we expect that children will use a different learning strategy than adults, particularly by focusing more strongly on the outcome of the most recent interaction. Research from the domain of developmental psychology suggests that trust relationships with peers already exist at a young age, but are initially very fragile and become more stable over the years. These studies on social relationships show that children typically understand the norm of direct reciprocity by 5 to 6 years of

age (Berndt, 1977; Youniss, 1980). Next, between ages 8 and 11, children still primarily base their estimation of trustworthiness on the most recent salient behavior of the other. Only at a later age trust is increasingly based on consistent patterns of behavior over time (Rotenberg & Pilipenko 1983-1984). At the latest stage of development, starting around early adolescence (12- 13 years of age) and lasting until late adolescence, friendships become increasingly stable and resistant to violations of trust (Kahn & Turiel, 1998).

In sum, these results based on questionnaires and self-reports, suggest that children will focus more on the most recent interaction when deciding what to in the next encounter thus play more strict tit-for-tat like than adults. Furthermore, it is well known that children are less capable to regulate their emotions in social situations than adults (Eisenberg, 2000). And because emotion regulation is thought to develop until at least late adolescence/young adulthood (Blakemore, 2008, Casey et al., 2008), we expect that children will be particularly sensitive to trust violations compared to adults. As a result, children will show higher levels of negative reciprocity; if in the previous round a player decided not to share, even if that player predominantly decided to share in the last few rounds. We therefore expect that children will often decide not to trust in the subsequent round, whereas adults might be more forgiving.

Trustworthiness, Anger and Punishment

Although the main goal of the current experiment was to investigate the ontogeny of relation-specific learning of trust, the current set-up also allowed us to further address the relation between emotional reactions to violated trust, and subsequent (costly) punishment of the violator. Unreciprocated trust and non-cooperative behavior in general are known to cause personal distress and, in particular, anger towards the non-cooperator (Pillutla & Murnighan, 1996; Stouten, De Cremer & van Dijk, 2009; Seip, van Dijk & Rotteveel, 2009). In addition, it is often assumed that the anger towards uncooperative norm violators, in this case of the norm of reciprocity (Gouldner, 1960), may motivate people to punish the perpetrator (Pillutla & Murnighan, 1996), even when this punishment is costly (Fehr, 2002; Fehr & Fischbacher 2004). Although there is some evidence for a causal relation between anger and punishment, this is not yet well established (see Seip et al., 2009). Furthermore, to our knowledge there are currently no studies that have investigated the relation between negative affect and costly punishment in developmental populations.

Based on two different strands of evidence we expect that children will show more negative affect towards norm violations than adults, and subsequently also higher levels of punishment. First of all, because children are less capable to regulate their emotions in social situations than adults (Eisenberg, 2000;

Steinberg, 2008), we expect that the anger evoked by unfair behavior will be higher for children than for adults. The increased anger could in turn lead to an increase in the level of punishment. This hypothesis is supported by studies that show that reduced self-regulation is strongly related to increased levels of reactive aggression (Conner, Steingard, Cunningham, Anderson & Melloni, 2004; Winstok et al., 2009). Reactive aggression is a particular form aggressive behavior that is evoked by perceived threat or provocation (Dodge & Coie, 1987), in this experiment the violation of trust.

Second, circumstantial evidence for our hypothesis that children will punish non-cooperators more than adults comes from developmental studies with the Ultimatum Game. In these studies participants are offered a split of a certain amount of money between themselves and another player. The results of these studies show that children reject unfair offers (unequal splits in advantage of the other player) more often than adults do (Murnighan & Saxon, 1998; Sutter, 2007; Güroğlu, van den Bos & Crone, 2009). Such rejections have been interpreted as means to punish, as they directly reduce the outcomes of the proposer.

In sum, there is some evidence for higher levels of anger and punishment in children compared to adults. However, no previous study investigated the relation between these two concepts in developmental populations. To investigate the relation between negative affect and costly punishment, we will measure the participants'feelings of anger towards the other players and their use of (costly) punishment (cf. Fehr, 2002) after they have finished the TG.

6.2 Method

6.2.1 Participants

Our sample included 60 participants (30 male, 30 female) divided over three age groups; late childhood (*M* age = 11.33, SD = 0.48, 9 male, 9 female), mid adolescence (*M* age = 16.24, SD = 0.91, 13 male, 8 female) and young adulthood (*M* age = 21.06, SD = 2.27, 8 male, 13 female). Chi-square analyses indicated that gender distributions did not differ significantly between age groups, $\chi^2(3) = 5.69$, $p = .078$. Children and adolescents were recruited by contacting local schools. Child and adolescent participants were selected with the help of their teachers (children with learning or psychiatric disorders were excluded); informed consent was obtained from a primary caregiver. Adults were recruited at the university.

6.2.2 Simultaneous Trust Game

To study how participants learn who to trust or distrust in a Trust Game setting we employed the Simultaneous Trust Game (STG) with repeated interactions. In the STG (Figure 6.1) the participants played multiple Trust Games in which both players simultaneously had to make their decision. Participants played the STG with three different players. At the start of each round the screen displayed the first name and photograph of the other player, who was always matched for age and gender. Next, the participant saw the complete decision tree and had to choose from two options: to trust or not to trust. If the participant decided not to trust, the coins were divided evenly, one euro each, between the players. If the participants decided to trust the other player the total money in the game was tripled in value (new total three Euros). When the other player had decided to reciprocate the 3 Euros was again divided evenly, one euro and fifty cents each, between the players. However, if the other player decided to defect she would take all the three Euros and leave the participant with nothing. The pay-off structure of the game was the same for every round (see Figure 6.1).

In the STG both players independently made their decision before the decision of the other is revealed, and in the end both decisions were always revealed. Thus before the decision of the participant to trust or not is revealed, the other player already had to decide if she would share or take all the money *if* she was trusted by the participant. Because the choice of the other player was always revealed, it was possible for the participants to learn what the trust outcome would have been even if they decided not to trust the other. Thus, if the participant chose not to trust that could result in two counterfactual outcomes; either the second player would have reciprocated trust or she would have defected trust and taken all the money. As a result, all participants (even those that never trusted) gained exactly the same information about the other players' decisions to share or not during the experiment.

The participants were told that the other player made his or her decisions through an internet connection but in reality the choice was made by the computer program and was displayed after a variable delay of 2-4 seconds. The presentation of this decision of the other player was displayed with an arrow by the outcome of choice. Blue arrows indicated a real outcome following a trust decision; grey arrows indicated a counterfactual outcome following a no trust decision. The presentation of the outcome of the trial was displayed for 3 seconds.

Participants were informed that during the experiment they were playing against three other unknown players. However, they actually played with computer-simulated agents with different pre-programmed strategies. The players were programmed with different percentages of sharing choices

(Trustworthy: 80%, Neutral: 50% and Untrustworthy: 20%). To represent the other players we used photographs of participants of the same age and gender. Prior to the experiment, the pictures where judged independently by 8 students on trustworthiness. Based on those judgments the most neutral faces on the trust dimension were selected for the experiment. To ensure that the individual characteristics of the faces did not bias trusting behavior we randomized the different faces over the different strategies. In total, the task consisted of 30 interactions with the 3 computer players. Consequently, for each participant the task consisted of 90 rounds in total. In each round the computer randomly picked one of the three other players, and the total number of rounds was unknown to the participants. The experiment was self-paced and took about 15 minutes complete.

Finally, the participants were told that the money they earned in the game would be exchanged for real money they would receive at the end of the experiment. We did not mention what the exact exchange rate between game and real money would be, but emphasized that the more money they earned the higher their real pay-off would be. Furthermore, the participants were told that their personal income would be revealed only when all other participants finished the experiment.

6.2.3 Post-Game Questionnaire

Right after the last round of the STG the participants filled in a computer based questionnaire. This questionnaire was not mentioned to the participants before they played the STG in order not influence their behavior in the game. We asked 3 questions regarding the frequency of sharing decisions of the other, level of trustworthiness and feelings of anger. The first three questions could be answered on a 5 point scale, ranging from not at all to very (often). We asked the participants to indicate their estimations of the frequency of sharing and levels of trustworthiness of the other players in order to check whether the different age groups have a comparable perception of how the other players behaved during the game, and how perceived behavior of the others is related to perceived trustworthiness of those players.

Finally, the participants had the opportunity to punish the other players by reducing some of their earnings. However, this punishment was costly; for each coin (€ 50 cents) paid by the participants the other player would lose 3 coins (€ 1.50, cf. Fehr, 2002). For each of the other players the participant could choose to pay an amount between zero and two Euros in increments of € 50 cents. The of the presentation of the three other players was randomized across participants.

6.2.4 Procedure

Child and adolescent participants were individually tested at their school in a quiet room and adult participants were tested in a laboratory, using a standard desktop computer or a laptop. Before the experiment started all participants received verbal instructions and had to fill out a questionnaire to test whether they understood the structure of the game. Subsequently, they played 10 practice rounds to get familiar with the interface. In case participants made mistakes in the questionnaire, the experimenter personally went over the questions with the participant to verify instructions were understood and if they were not correct they would go through another set of practice rounds until the task was understood fully.

6.2.5 Instructions

All participants got their picture taken a week before they participated in the experiment, and were told their picture would be shown the other players they interacted with in the experiment that would follow. The participants were instructed that they were going to play an interactive game with two other players with whom they were connected via the internet. Furthermore, they were told that at the end of the experiment the computer would determine the pay-off for all players. It was emphasized that therefore their decisions had consequences for the pay-off of themselves *and* others. The total duration of the experiment was approximately 35 minutes. Last, when all participants had completed the experiment, all participants were paid and debriefed about the actual set-up.

6.3 Results

First we tested whether the different age groups differed in their perceptions of frequency of sharing and trustworthiness. Next we investigated how participants of different ages learned who to trust and distrust, and the relation between age, anger and punishment.

6.3.1 Manipulation check

To check whether there were age differences in the perception of the frequency of sharing decisions of the three types of players we performed ANOVA with frequency of reciprocal choices as dependent variable, type of player as within-subjects variable and age group as between-subjects factor. These analyses revealed a main effect of Type ($F(2,58) = 163.20$, $p < .001$), but no effect of Age ($F(2,58) < 1$, $p = .67$). Participants of all age groups recognized that the three players differed significantly in their frequency of sharing decisions (see Table 6.1), and frequency estimations did not differ between age

groups. These results are important because they show that age differences in punishing behavior or emotions are not due to different perceptions of the strategies of the other players.

To investigate whether the different strategies of the other players were correctly recognized as differences in trustworthiness we performed a similar ANOVA with trustworthiness as dependent variable. As expected, these analyses revealed a main effect of Type ($F(2,58) = 138.22$, $p < .001$), but no effect of Age ($F(2,58) < 1$, $p = .51$). That is, participants of all age groups perceived the three players differing significantly in their trustworthiness (see Table 6.1), but importantly these estimations did not differ between age groups.

Table 6.1: Average levels of Frequency estimation and Trustworthiness collapsed over all age groups.

	Trustworthy	Neutral	Untrustworthy
Frequency	4.46	2.93	1.63
Trust	4.03	2.72	1.70

6.3.2 Generalized Trust – the first move

As expected, our data show that 11 year olds that made fewer trust decisions (M = 27%) in the first round relative to the 16 year olds (M= 47%) and the 22-year-olds (M = 70%) who trusted the most (see Figure 6.2). Indeed, a logistic regression with first choice as dependent variable and age group as covariate revealed that with increasing age participants showed significantly more trust in the first round ($\beta = .90$, $p < .01$, see Figure 6.2).

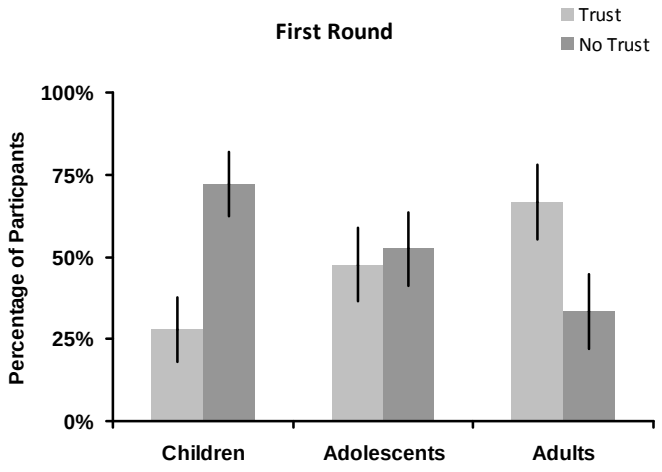


Figure 6.2: The percentage of participants in each group that decided to either trust and not to trust in the very first round of the experiment. Error bars represent standard error.

6.3.3 The relation-specific changes in trust

Next, we were interested in how trust relations changed over time based on the behavior of the other player, and whether there were age differences in these developing patterns of trust. To investigate the relation-specific changes in trust over time we divided the experiment in three equal blocks (begin, middle, end). We performed a repeated measures ANOVA with Type of player (trustworthy, neutral, untrustworthy) and Time (begin, middle, end) as within-participants factors and Age as between-participants factor for the percentage of trust choices.

As expected, this analysis yielded a main effect of Type of player on trust, $F(2,58)= 128.03$, $p < .001$, which was qualified by a significant Type x Time interaction ($F(4,58)= 8.98$, $p < .001$); over time participants showed *increasing* trust for the trustworthy player and *decreasing* trust for the untrustworthy player (see Figure 6.3). Moreover, our analyses also revealed an Type x Time x Age interaction ($F(4,58)= 13.14$, $p < .005$). This indicates that there are age differences in relation-specific changes in trust, as can be seen in Figure 6.3.

To further interpret these age differences we performed separate Type x Time ANOVAs for each age group. These analyses revealed that in all age groups there was a significant difference in the amount of trust in each of three players (i.e., a main effect of type, all p's $< .001$). Furthermore, for adults and adolescents ($F(4,18)= 16.11$, $p < .001$ and $F(4,19)= 5.74$, $p < .005$ respectively) but not for the children ($F(4,18)= 2.67$, $p = .08$) we observed a significant Type x Time interaction. The pattern of the children differs from the other age groups by showing no significant change in strategy over time, whereas adults and adolescents started to trust the trustworthy player more, and the untrustworthy player less, over time (see Figure 6.3).

6.3.4 Tit for tat?

To investigate differences in strategy use during the game we analyzed sequential effects. We analyzed whether the choice of the other player to reciprocate or defect in the previous round, regardless of whether the outcome was real or counterfactual (i.e., following trust or no trust), influenced the participants' decision to trust in the next round with the same player. If the participants followed a reciprocal tit-for-tat like strategy they would decide to trust if the other player had decided to share in the previous round (positive reciprocity), and decide not to trust when the other player had decided to keep all the money in the previous round (negative reciprocity). For these analyses we therefore calculated the percentage of tit-for-tat choices the participants made and compared these percentages for each age group with a univariate ANOVA.

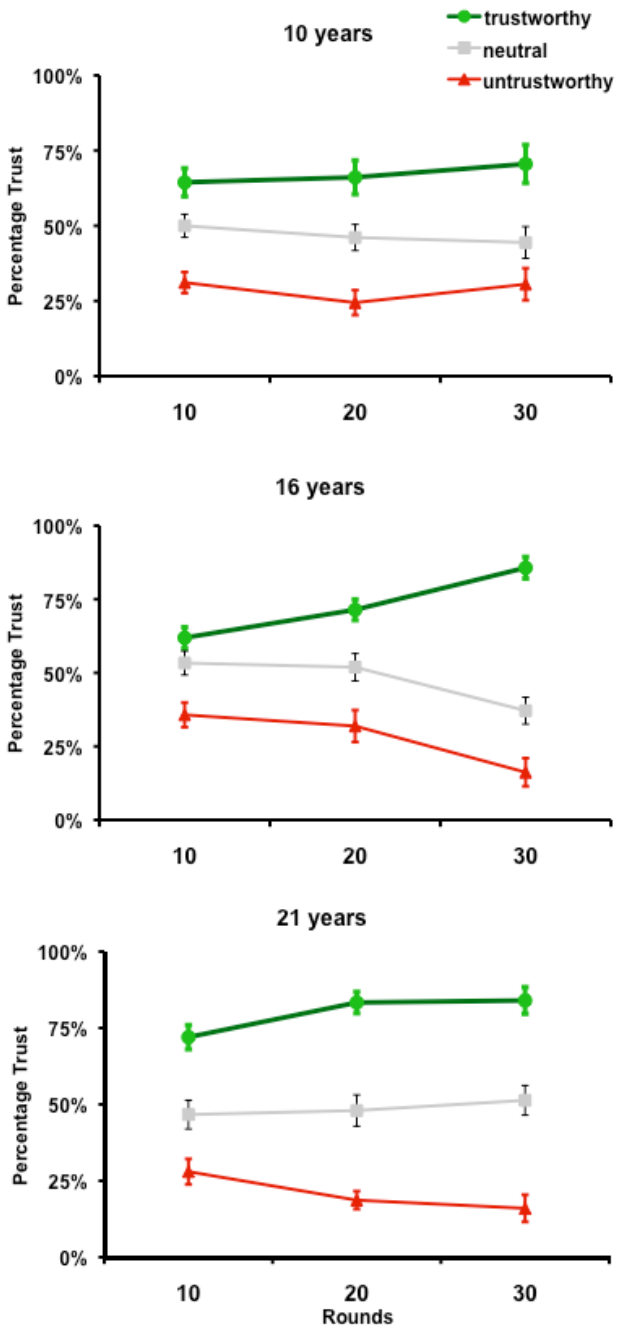


Figure 6.3: The mean percentage of trust decisions per block of 10 trials for both computer players, error bars represent standard error.

This analyses revealed that all age groups applied a tit-for-tat strategy to a certain extent (all groups are well above 50% see Figure 6.4), and that there was a main effect of Age ($F(1,58)= 6.41$, $p < .001$). Tukey's b tests for post hoc comparisons with an alpha of .05 showed that children (89 %) applied the tit-for-tat strategy more often than adults and adolescents (74% and 71% respectively), who did not significantly differ in their strategies.

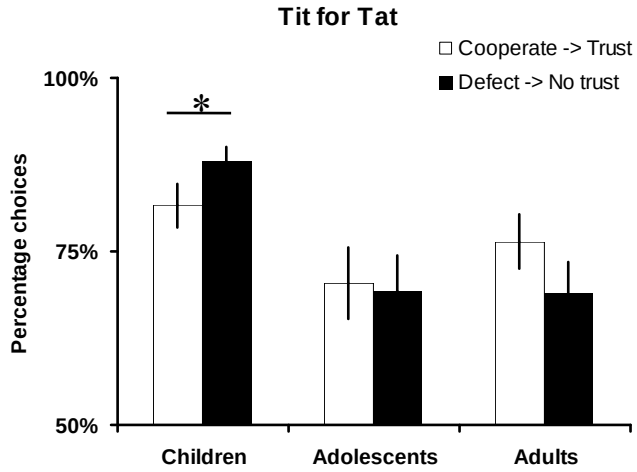


Figure 6.4: The mean percentage of negative and positive tit-for-tat decisions for all age groups.

6.3.5 Positive vs. Negative Reciprocity

A tit-for-tat strategy is characterized by both positive and negative reciprocity. However, because psychologically positive and negative reciprocity may represent distinct processes we also analyzed them separately. To investigate possible differences in tit-for-tat choices after either a share or keep choice of the other we preformed an ANOVA with Type of tit-for-tat (positive vs. negative) as within-participants factors and Age as between-participants factor. This analysis revealed further group differences in strategy use (Type of tit-for-tat x Age interaction, $F(2,58)= 4.67$, $p < .01$). Post hoc paired t-tests per age group shows that children showed higher levels of negative than positive reciprocity ($t(1,17) = -2.46$, $p < .025$), whereas the other groups did not show such a difference (both p's > .3, see Figure 6.4).

Finally, based on the changing patterns of trust behavior over time we also investigated whether the number of tit-for-tat choices changed over time. For

this purpose we calculated the percentage of tit-for-tat choices in each block (begin, middle, end) with age group as between subjects factor (see Figure 6.5). This analysis resulted in an Age x Time interaction, $F(2,58)= 3.67$, $p < .04$. Subsequent ANOVAs per age group revealed a significant effect of Time on tit-for-tat strategy ($F(2,18)= 5.21$, $p < .02$) for adults, but not for the children and adolescents (both p's >.1). As can be seen in Figure 6.5, adults showed an increase in tit-for-tat choices between the beginning and middle period and which then stayed on the same level, whereas the children remained at a stable level of tit-for-tat choices from the beginning until the end.

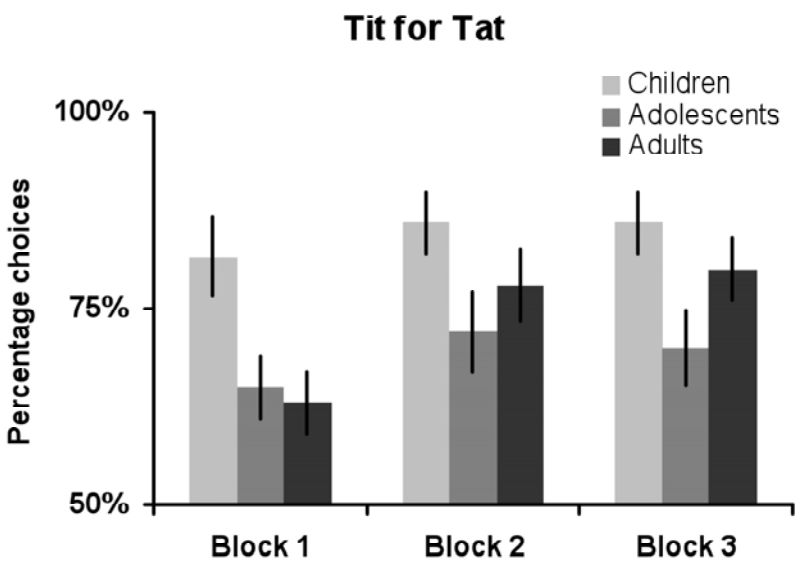


Figure 6.5: The mean percentage of tit-for-tat decisions per block of 10 trials for all age groups. In both graphs error bars represent standard error.

6.3.6 Anger and punishment

Next, we investigated how the different strategies used by the computer players elicited feelings of anger and subsequent punishment. We performed an ANOVA with anger as dependent variable, Type of player as within-subjects variable and age group as between-subjects factor. These analyses revealed a main effect of Type ($F(2,58) = 12.70$, $p < .001$), and a main effect of Age ($F(2,58) = 12.69$, $p < .001$). That is, participants of all age groups showed more anger to the least trustworthy person; but the younger participants also showed more anger than the older participants (see Figure 6.6).

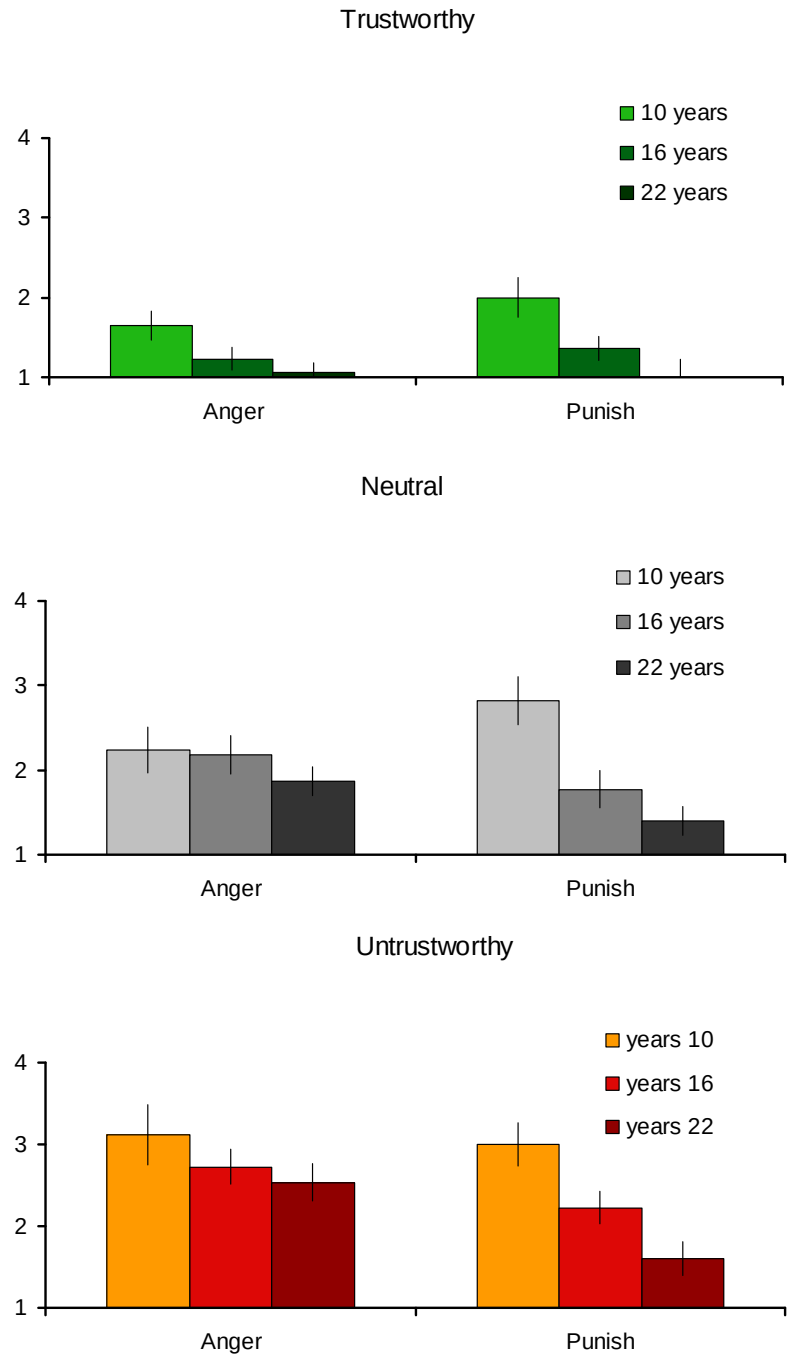


Figure 6.6: Post-Game Questionnaire results for anger and punishment.

The analysis of punishment behavior revealed a similar pattern to that of anger; participants of all age groups punished the least trustworthy person the most (main effect Type, $F(2,58) = 16.12$, $p < .001$), and with age there was a general decrease in the amount of punishment given (main effect Age, $F(2,58) = 5.08$, $p < .03$)

Finally, we were interested in the relation between reported levels of anger and subsequent size of punishment. As expected, there was a significant correlation between anger and punishment for all age groups ($r = .54$, $p < .01$, $r = .53$, $p < .01$ and $r = .42$, $p < .03$ for children, adolescents and adults). The correlation between anger and punishment suggests that the increase in punishment with age is a result of increased anger with age. To further investigate the relation between age and increased anger and punishment we have performed mediation analyses. According to Baron and Kenny (1986), mediation can be established by demonstrating that (a) there is a direct effect of the independent variable (i.e., age) on the dependent variable (i.e., punishment), (b) there is a significant effect of the independent variable on the proposed mediator (i.e., anger), (c) the proposed mediator is correlated with the dependent variable after controlling for the independent variable, and (d) the effect of the independent variable on the dependent variable drops significantly when the mediator is included in a simultaneous regression (Baron & Kenny, 1986). As can be seen in Figure 6.7, all of the Baron and Kenny requirements are met. First, there is a significant effect of age on punishment ($\beta = -.43$), $t(60) = 2.17$, $p < .02$, and on the proposed mediator, anger ($\beta = -.31$), $t(60) = 2.9$, $p < .005$. Second, anger was correlated with punishment when controlling for age ($\beta = .41$), $t(60) = 2.86$, $p < .003$. Finally, the direct effect of sharing frequency was no longer significant ($\beta = -.21$), $t(60) = 1.13$, $p = .26$, when controlling for anger. In summary, the influence of age on punishment was completely mediated by feelings of anger towards the other player (Sobel $z = 3.91$, $p < .01$).

Figure 6.7*:* Mediation analyses of age, anger and punishment. ** $p < .01$, * $p < .05$

Mediation also can be demonstrated by showing that the indirect effect (i.e., the path through the mediator) is significantly different from zero. The indirect effect is the product of two regression coefficients; specifically, the product of the regression weight linking the independent variable to the mediator (denoted *a*) and weight linking the mediator to the dependent variable (denoted *b*). Shrout and Bolger (2002) suggest that a formal test of mediation be conducted using a bootstrapping technique that involves computing confidence intervals around the product term (*a*b*). If zero falls outside of this 95% confidence internal, the indirect effect is significant and mediation can be said to have occurred. To implement this approach, we used SPSS syntax provided by Preacher and Hayes (2004) using 10,000 iterations. This approach provided results consistent with the mediation analyses described earlier. Specifically, zero fell outside our 95% confidence interval around the indirect effect, which ranged from .17 to .54. These results provide converging evidence that anger mediates the effects of frequency of sharing on punishment.

6.4 Discussion

Despite strong evidence for the benefits of trust for social development and society (e.g. Fukuyama, 1995; Bernath & Feshbach, 1995), it is less well known how people learn to trust or distrust persons and anonymous institutions. In this article, we combined insights from social psychology and developmental psychology, and used the STG to study the relation-specific changes in trust or distrust in three age groups. To our knowledge, no study to date has investigated age differences in learning who to trust and costly punishment of trust violations. This study had two main goals: (1) To examine the development of trust relationships between late childhood and young adulthood, and (2) To examine the developmental trajectory of emotions evoked by non-cooperative behavior of others, and to what extent these emotions may lead to altruistic punishment. To this end, the discussion is organized according to these main goals.

Changes of trust

As noted, previous research with the Trust Game has paid some attention to trust in children, but almost without exception these studies involved adults only. The decisions of adults in the current study resemble the pattern typically seen in these behavioral experiments. That is, adult participants often chose to trust in the first round, indicating that they expected others to reciprocate (e.g. Berg et al., 1995; Dufwenberg & Gneezy, 2000; McCabe et al., 2001). However, there were important age related changes in first move. As expected, children showed a low level of general, trust; most of them started with not

trusting the other. In addition, consistent with previous studies (Sutter & Kocher 2008; van den Bos et al., 2009) our analyses revealed that this general trust increases with age.

Next, we investigated the 'relation-specific' trust, by analyzing how levels of trust changed over time based on the interactions with the three different players. As expected, the strategy of the other player influenced the percentage of trust choices; over time adults learned who to trust and who to distrust. Furthermore, our analyses revealed that they applied a tit-for-tat like strategy, and importantly their strategy also changed over time. In the initial phase of the experiment, adults quickly learned the level of trustworthiness of the other players and adapted their behavior accordingly. At the moment they learned who to (dis)trust, their strategy changed to a stable tit-for-tat like pattern. Together, these results suggest that for adults trust decisions are initially based on a fairly high level of general trust and then are quickly adapted to the level trustworthiness of the player they interacted with. Our current study extends these findings by investigating how children and adolescents learned who to trust.

Although the low level of general trust displayed by children in the first trial is consistent with previous studies, the following question remained: how would children and adolescents learn to trust or distrust another player? Our analyses of the relation-specific changes in trust revealed that participants of all ages were able to learn to trust a certain player, and importantly also learn not to trust another player. Indeed, both children and adolescents ended with high levels of trust for the trustworthy player and low levels of trust for the untrustworthy player. Interestingly, there were also age differences in strategies. As expected, the children used the strictest form of tit-for-tat strategy compared to the other age groups, and they did not show any changes in their strategy during the experiment. As a result of this strategy, the children displayed the same level of trust from beginning (excluding the first move) until the end, whereas we observed significant changes in levels of trust for both adolescents and adults. Finally, further analyses revealed that children differed from adults and adolescents especially in showing higher levels negative reciprocity.

Why would children use a more direct tit-for-tat like strategy, in particular after trust violations, compared to adults and adolescents? The tit-for-tat strategy could be partly due to the higher levels of reactive aggression displayed by younger children (Conner et al., 2004); they keep reacting strongly to violations of trust in the previous round regardless of their indication of trust behavior in the past. This would also explain that the children particularly showed higher levels of negative reciprocity compared to the other age groups. Another possible, but tentative, explanation for the high level of tit-for-tat

choices for the youngest participants is that tit-for-tat is a strategy that requires very little from working memory (Milinski & Wedekind, 1998). Given that children still have an underdeveloped working memory capacity (e.g. Crone et al., 2006), they are more likely to use a strategy that is less memory demanding. This explanation is consistent with studies that have shown that adults will start playing more direct tit-for-tat when their memory load is occupied by another task (Milinski & Wedekind, 1998). However, because age did not significantly affect the participants' estimations of the total frequency of sharing decisions, we consider this explanation less likely. Nonetheless, it would be interesting for future research to investigate the possible relation between age related differences in working memory capacity and strategy differences in social interactions.

Anger and Punishment

Next, we investigated participants' emotional reactions to trust violations and levels of costly punishment. As expected, the three players evoked different levels of both anger and punishment. Participants of all age groups were most angry at the player that violated trust the most, and punished accordingly. This pattern of behavior is consistent with several previous studies that investigated the relation between anger and costly punishment (see Seip et al., 2009). However, there were also large differences in levels of anger between age groups. Although all participants displayed more anger towards those players that violated trust the most, children showed more anger than adolescents, and adolescents more than adults. Additionally, we found that the younger participants punished more than the older participants.

In contrast to children, adults showed virtually no anger towards, and did not punish the least untrustworthy player, even though that player kept the money 20% of the time. So, although that player displayed some trust violations, these occasional violations did not seem to anger the adult participants, and it did not induce them to punish. This finding is in line with previous work that showed that positive peer relations become more stable and resistant to violations of trust (Kahn & Turiel, 1998). Future studies could further test the hypotheses of increasing stability with age by studying how participants adjust their behavior when their trustworthy player unexpectedly changes into an untrustworthy one. Based on our hypothesis above we would expect that children will almost immediately adjust, applying their strict tit-for-tat strategy, whereas adults would take longer to adjust because they also take the longer history of interactions into account.

This leaves the question why children and adolescents reacted angrier than adults? One possible explanation is that they are less able to regulate the anger

evoked by violations of trust. This interpretation is in line with studies that show that age related increases in emotion regulation are strongly related to lower levels of reactive aggression (Conner et al., 2004; Winstok, 2009).

Another explanation for the higher levels of anger in children is that their affective reaction to social interaction is based on a different perception of the intentions of the other players. Previous developmental studies have suggested that the increased skill of perspective taking, the ability to reason about the others' intentions, significantly changes social behavior in one-shot Trust (van den Bos et al., 2009) and Ultimatum Games (Sutter, 2007; Guroglu et al., 2009). These studies suggest that an age related increase in perspective taking may lead to increased trust and a decrease in rejection rates. Furthermore, Mohr and colleagues (1999) showed that increased anger after provocation (i.e. violation of trust) is significantly related to a decreased capability of perspective taking (see also Eisenberg et al., 2006). Taken together, this suggests that a possibly more negative perception of the others' intentions by the younger participants could have led to more anger and subsequently more punishment after the violation of trust. Given that all age groups had similar perceptions of the trustworthiness of the three players, the current results favor the explanation of differences in emotion regulation over perspective taking. Future studies may focus on disentangling the effects of perspective taking and emotion regulation on the increased negative affect in developmental populations, which will further our understanding of these processes in social decision-making.

Conclusion

The current finings revealed the importance of several psychological processes involved in learning who to trust. A comparison of age differences of behavior in the STG indicates that, besides a general increase of generalized trust, relation-specific trust changes with age. In particular, children appeared to apply a stricter tit-for-tat like strategy than adults and adolescents, and seem especially more sensitive to violations of trust. Additionally, the results show that with increasing age the amount of both anger and punishment decrease, and that age differences in trust were fully mediated by feelings of anger. Together these results indicate that the stability of adult trust relationships might be the results of an age related reduction of negative affect and negative reciprocity towards the violations of trust. Moreover, the current findings demonstrate how the combination and integration of social psychological and developmental insights may contribute to understanding of how we learn to trust (and distrust) others.

7. Better than expected or as bad as you thought? The neurocognitive development of probabilistic feedback processing

Abstract

Learning from feedback lies at the foundation of adaptive behavior. Two prior neuroimaging studies have suggested that there are qualitative differences in how children and adults use feedback by demonstrating that dorsolateral prefrontal cortex (DLPFC) and parietal cortex were more active after negative feedback for adults, but after positive feedback for children. In the current study we used functional magnetic resonance imaging (fMRI) to test whether this difference is related to valence or informative value of the feedback by examining neural responses to negative and positive feedback while applying probabilistic rules. In total, 67 healthy volunteers between ages 8 and 22 participated in the study (8–11 years, n = 18; 13–16 years, n = 27; 18–22 years, n = 22). Behavioral comparisons showed that all participants were able to learn probabilistic rules equally well. DLPFC and dorsal anterior cingulate cortex were more active in younger children following positive feedback and in adults following negative feedback, but only when exploring alternative rules, not when applying the most advantageous rules. These findings suggest that developmental differences in neural responses to feedback are not related to valence *per se*, but that there is an age related change in processing learning signals with different informative value.

7.1 Introduction

Learning to correctly adapt your behavior in a changing environment is an essential feature of human cognition and has been studied extensively over the past decades (for reviews, see Ridderinkhof and van den Wildenberg, 2005; Rushworth and Behrens, 2008). When adapting behavior, individuals often make use of feedback signals, which can be positive, encouraging the continuation of behavior, or negative, discouraging the continuation of behavior and signaling the need for adjustment. Prior studies have indicated that adaptive learning based on feedback signals undergoes pronounced developmental improvements between late childhood and early adulthood, as is evident from tasks in which participants need to switch between multiple rules (Crone and van der Molen, 2004; Somsen, 2007) or in which they need to infer sorting rules based on positive and negative signals (van Duijvenvoorde et al., 2008).

Early developmental improvements in adaptive behavior are observed when feedback has a direct mapping to deterministic rules (Somsen, 2007), however, when the feedback is probabilistic, changes in adaptive learning are observed until late adolescence (Hooper et al., 2004). In these situations, individuals must learn the statistical regularities between actions and outcomes, and use that information to interpret current feedback signals (see also Rangel et al., 2008). Feedback which is not directly mapped to behavior is often more complex because it requires individuals to attend to long term consequences and override the tendency to respond directly to local environmental change.

Neuroimaging studies have shown that regions previously associated with cognitive control and response selection (Miller and Cohen, 2001; Toni et al., 2002) are also active when adults receive negative performance feedback, including the dorsal anterior cingulate cortex (dACC) and the dorsolateral prefrontal cortex (DLPFC) (Klein et al., 2007; Taylor et al., 2007). The dACC is thought to monitor action outcome regularities and is important for signaling adjustment (Botvinick et al., 2001; Yeung et al., 2004). In addition, the dACC may exercise behavioral control via the engagement of the DLPFC (Kerns et al., 2004; Zanolie et al., 2008), which in turn is important for trial-to-trial adjustments of behavior (Dosenbach et al., 2008). Similar to the DLPFC, the parietal cortex is also involved in feedback processing, in particular negative feedback (Crone et al., 2008; van Duijvenvoorde et al., 2008). Finally, these regions are thought to work in close concert with the basal ganglia, specifically the caudate nucleus, which is thought to be engaged when learning action-outcome regularities (for a review see Cools, 2008).

In two prior developmental studies we have identified the developmental time course of these regions during adaptive feedback processing. In the first study (Crone et al., 2008), participants were instructed to infer rules based on positive and negative feedback which could change without warning. Following Somsen (2007), we were interested in the way children, adolescents, and adults processed negative feedback indicating a rule shift. As anticipated, adults engaged DLPFC, dACC, and the parietal cortex when processing negative feedback indicating a rule shift. A similar pattern was observed in 14- to 15-year-old adolescents, but 8- to 11-year-old children engaged these regions less following negative feedback in comparison to positive feedback or a low-level fixation baseline. In the second study (van Duijvenvoorde et al., 2008), participants were instructed to guess a correct rule. Because there were two possible rules, there was a 50% chances of receiving positive feedback, and therefore both feedback signals (negative and positive) were similarly salient and probable. Again, adults engaged DLPFC, dACC, and the parietal cortex following negative feedback, but in this study 8-year-old children engaged

DLPFC and the parietal cortex more following positive feedback relative to negative feedback. The developmental trajectory of the dACC followed a different pattern, as it slowly emerged in response to negative feedback at the age of 12, but it was not more active following negative compared to positive feedback at a younger age (see also Velanova et al., 2008). Although the caudate nucleus was involved in these tasks, these studies revealed that there were no developmental differences in activation patterns.

Together, these findings indicate that the possible meaning of positive and negative feedback signals, and the role of the associated neural circuits, changes during development. However, prior studies could not dissociate between neural activation as a result of valence versus informative value, given that negative feedback always signaled response adjustment and therefore had different informative value than positive feedback. Thus, it remains to be determined how the involvement of DLPFC and the parietal cortex is dependent on valence versus informative value of the feedback.

Prior research suggests that differences in positive and negative feedback adjustment are the result of differences in attention regulation (Somsen, 2007). Following this hypothesis, it is argued that children are less able to update the relevant feedback information and therefore they are less flexible in selecting alternative actions. We therefore reasoned that the brain regions implicated in prior feedback studies may be sensitive to the informative value of feedback, and that activation in these brain regions is indicative of feedback attendance. Furthermore, we predicted that attention to feedback may also underlie the developmental differences in brain activation. We hypothesized that DLPFC and parietal cortex would be more active following positive feedback in children and following negative feedback in adults, but only when the feedback has informative value for learning and response adjustment. Thus, we sought to test how neural responses are sensitive to informative value for learning versus valence of feedback, and the developmental trajectory of feedback processing.

We reasoned that feedback valence versus informative value could be disentangled after participants learned probabilistic feedback rules. In the probabilistic learning paradigm, participants need to learn from positive and negative feedback under different levels of probability, and therefore not all positive feedback signals response continuation and not all negative feedback signals response adjustment. The probabilistic learning (i.e., trial-and-error) task employed in this study was based on a prior study by Frank et al. (2004), but was simplified for use with children. In our version of the probabilistic learning task, two different stimulus pairs (AB or CD) were presented in random order, and participants had to learn over trials that one stimulus was more likely to result in positive feedback (70–80%) (see Figure 7.1). Over the course of the

experiment participants had to learn the statistical regularities and thus had to learn to choose the stimuli with a high probability of positive feedback (A and C) more often than those with a low probability of positive feedback (B and D).

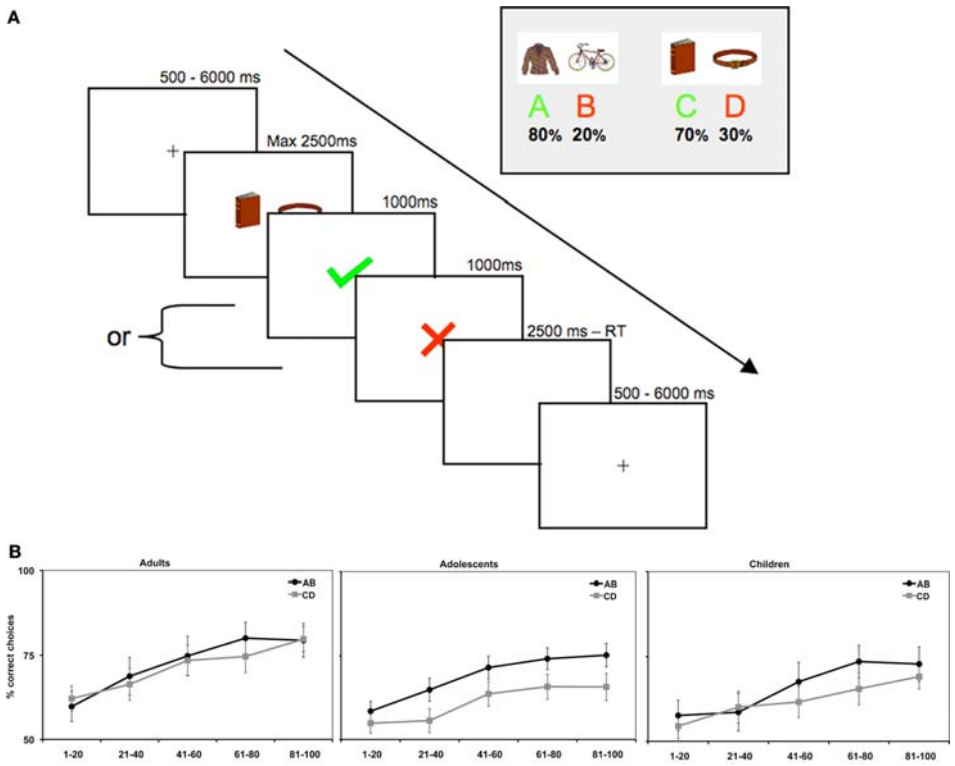


Figure 7.1 (A) At the beginning of each trial a centrally located cue was presented with a jittered interval between 500 and 6000 ms, followed by a combined presentation of a stimulus pair and a response window of max. 2500 ms, after which feedback was presented for 1000 ms. After the feedback a short filler was presented, in the form of a blank screen, in order to compensate for different reaction times between trials and between participants (filler duration = 2500 ms – reaction time). (B) Average accuracy on AB and CD trials per age group.

When participants have gained knowledge of the statistical regularities, they were expected to more often apply the correct rule. Notably, in probabilistic learning tasks individuals generally do not consistently apply the correct rule but show matching behavior; i.e., they choose the correct stimulus with a frequency that is proportional to the probability of positive feedback associated with that stimulus (Estes, 1961; Herrnstein, 1961; Shanks et al., 2002; Frank and Kong, 2008). Thus, we anticipated that participants would apply the correct rule (in this study, choosing the high probability stimuli A and C) more often,

but we also anticipated that they would remain exploring the alternative rule (choosing the low probability stimuli B and D). Therefore, this paradigm allowed us to investigate the processing of positive and negative feedback that carries different informative value. In particular, receiving negative feedback when choosing the correct rule should not be interpreted as a signal to switch to the alternative rule because the probability of positive feedback remains higher than for the alternative rule. In contrast, receiving negative feedback when choosing the alternative rule should lead to a switch to the correct rule. To be able to address the question how neural responses are sensitive to feedback signals in the context of learned rules, we only analyzed neural responses after participants had reached a learning plateau.

Based on prior studies, we expected that DLPFC and the parietal cortex would be sensitive to whether feedback signals required greater attention, and would contain greater informative value for performance adjustment on subsequent trials. Therefore, we expected that these regions would be engaged mostly after choosing the alternative rule (B or D), because this feedback contained learning signals for performance adjustment, independent of valence. We also examined the role of the dACC and the caudate as these regions have previously been implicated in feedback processing (Schultz, 2007; Cools, 2008; Rushworth and Behrens, 2008). We expected that the dACC would be most sensitive to negative feedback signals, particularly when indicating the need for behavioral adjustment (Kerns et al., 2004), whereas we expected that the caudate would be most sensitive to positive feedback which signals response continuation (Cools, 2008).

The second question concerned developmental differences in performance and neural activation. In prior research, developmental differences were observed between childhood and mid-adolescence, but differences between adolescence and adulthood remain unclear (Crone et al., 2008; van Duijvenvoorde et al., 2008). For this purpose, we compared behavioral and neural responses of three age groups; children (8–11 years), adolescents (13–16 years), and adults (18–22 years). Behaviorally, we predicted that differences in adaptive learning would be largest between childhood and adolescence, with refinement of learning between adolescence and adulthood (Luna and Sweeney, 2001; Crone and van der Molen, 2004; Somsen, 2007). In addition, we expected to find that these behavioral changes would be paralleled by changes in the areas involved in adaptive control (dACC, DLPFC, parietal cortex and caudate nucleus). For the fMRI analyses, we had three specific age related hypotheses based on prior studies. First, we expected an increase in differentiation in the dACC for positive and negative feedback processing with increasing age (van Duijvenvoorde et al., 2008; Velanova et al., 2008). Second, we expected an

attention-based shift in recruitment of DLPFC and the parietal cortex from positive to negative performance feedback with age. Third, we expected age differences in how learned probabilities would be associated with neural changes in feedback processing; in particular we predicted that feedback after exploring the alternative rule would be associated with developmental differences. Because of the children's putative focus on positive feedback, we expected that with increasing age there would be a decrease in activity related to processing positive feedback and an increase in activity related to processing negative feedback following selection of the alternative rule.

Finally, our paradigm allowed us to investigate age differences in adaptive behavior, that is, whether participants stay or shift on subsequent trials based on the received feedback. Besides behavioral analyses of sequential effects, we also employed exploratory sequential condition analyses to further understand the relation between neural activation and subsequent adjustment of behavior (see also Kerns et al., 2004).

7.2 Materials and Methods

7.2.1 Participants

Sixty-seven healthy right-handed paid volunteers (35 female, 32 male; ages 8–22 participated in the fMRI experiment. Age groups were based on adolescent development stage, resulting in three age groups: children (8- to 11-year-olds, n = 18; 9 female), mid-adolescents (13- to 16-year-olds, n = 27; 13 female) and young adults (18- to 22-year-olds, n = 22; 13 female). A chi square analysis indicated that the gender distribution was similar across age groups, $\chi^2(2) = 0.79$, $p = 0.67$. All participants reported normal or corrected-to-normal vision and participants or their caregivers indicated an absence of neurological or psychiatric impairments. Participants and their caregivers (for minors) gave informed consent for the study and all procedures were approved by the medical ethical committee of the Leiden University Medical Center. In accordance with Leiden University Medical Center policy, all anatomical scans were reviewed and cleared by the radiology department following each scan. No anomalous findings were reported.

7.2.2 Behavioral Assessment

Parents filled out the Child Behavior Check List (CBCL, Achenbach, 1991) for participants younger than 18 years, in order to screen for psychiatric conditions. All participants scored below clinical levels on all subscales of the CBCL, and had scores within 1 SD of the mean of a normative standardized sample.

Participants completed two subscales (similarities and block design) of either the Wechsler Adult Intelligence Scale (WAIS) or the Wechsler

Intelligence Scale for Children (WISC) in order to obtain an estimate of their intelligence quotient (Wechsler, 1991, 1997). There were no significant differences in estimated IQ scores between the different age groups, $F(2, 66) = 1.63$, $p = 0.20$ (see Table 7.1).

Table 7.1 Groups Measures

	IQ	RT (ms)	points	head motion Avg (mm)	Max (mm)
Adults	107 (2.4)	811(44)	118(3)	.08(.01)	1.56
Adolescents	108 (2.0)	773(39)	114(3)	.08(.01)	2.96
Children	111 (2.6)	804(42)	107(6)	.09(.01)	2.85

Displays means per age groups, standard errors between brackets. Final column represents the maximum head motion between two time points in each group

7.2.3 Task Procedure

The procedure for the probabilistic learning task (Frank et al., 2004) was as follows: The task consisted of two stimulus pairs (called AB and CD). The stimulus pairs consisted of pictures of everyday objects (e.g., a chair and a clock). Each trial started with the display of one of the two stimulus pairs and subsequently the participant had to choose one of the two stimuli (e.g., A or B), which were presented on the left or the right side of the screen. The stimulus pairs were presented in random order. Participants were instructed to choose either the left or the right stimulus by pressing a button with the index or middle finger of the right hand within a 2500 ms window, which was followed by a 1000 ms feedback display. The feedback display consisted of a green V-signal for positive feedback and a red cross for negative feedback. If no response was given within 2500 ms, the text "too slow" was presented on the screen. This occurred on less than 2% of the trials.

The feedback displayed was probabilistic. Choosing stimulus A led to positive feedback on 80% of AB trials, whereas choosing stimulus B led to positive feedback on 20% of these trials. The CD pair procedure was similar, but probability for positive feedback was lower; choosing stimulus C led to positive feedback on 70% of CD trials, whereas choosing stimulus D led to positive feedback on 30% in these trials. Thus, the correct choice in order to obtain most positive feedback was A or C, whereas the incorrect choice was B or D.

Participants were instructed to earn as many points as possible (as indicated by receiving a positive feedback signal), but were also informed that it would not be possible to receive positive feedback on every trial. Further, participants were informed that although stimuli sometimes appeared on the right side and sometimes on the left side, that laterality was an irrelevant dimension. After the

instructions and right before the scanning session, the participants played 40 practice rounds on a computer in a quiet laboratory to ensure proficiency on the task.

In total, the task in the scanner consisted of two blocks of 100 trials each: 50 AB trials and 50 CD trials per block. To ensure that participants had to learn a new mapping in both task blocks, the first and the second block consisted of different sets of pictures. The duration of each block was approximately 8.5 min. The stimuli were presented in pseudo-random order with a jittered interstimulus interval (min = 1000 ms, max = 6000 ms) optimized with OptSeq2 (Dale, 1999). During inter trial intervals, a central fixation cross was shown.

7.2.4 Data Acquisition

Participants were familiarized with the scanner environment on the day of the fMRI session through the use of a mock scanner, which simulated the sounds and environment of a real MRI scanner. Data were acquired using a 3.0T Philips Achieva scanner at the Leiden University Medical Center. Stimuli were projected onto a screen located at the head of the scanner bore and viewed by participants by means of a mirror mounted to the head coil assembly. First, a localizer scan was obtained for each participant. Subsequently, T2*-weighted Echo-Planar Images (EPI) (TR = 2.2 s, TE = 30 ms, 80 × 80 matrix, FOV = 220, 35 2.75 mm transverse slices with 0.28 mm gap) were obtained during two functional runs of 232 volumes each. The first two scans were discarded to allow for equilibration of T1 saturation effects. A high-resolution T1-weighted anatomical scan and a high-resolution T2-weighted matched-bandwidth high-resolution anatomical scan, with the same slice prescription as the EPIs, were obtained from each participant after the functional runs. Stimulus presentation and the timing of all stimuli and response events were acquired using E-Prime software. Head motion was restricted by using pillow and foam inserts that surrounded the head.

7.2.5 fMRI Data Analysis

Data were preprocessed using SPM5 (Wellcome Department of Cognitive Neurology, London). The functional time series were realigned to compensate for small head movements. Translational movement parameters never exceeded 1 voxel (<3 mm) in any direction for any subject or scan. There were no significant differences in movement parameters between age groups $F(2, 65) = 0.152$, $p = 0.85$, (see Table 7.1). Functional volumes were spatially smoothed using a 6 mm full-width half-maximum Gaussian kernel. Functional volumes were spatially normalized to EPI templates. The normalization algorithm used a 12 parameter affine transformation together with a nonlinear transformation

involving cosine basis functions and resampled the volumes to 3 mm cubic voxels. The MNI305 template was used for visualization and all results are reported in the MNI305 stereotaxic space (Cosoco et al., 1997), an approximation of Talairach space (Talairach and Tourneaux, 1988).

Statistical analyses were performed on individual participants' data using the general linear model in SPM5. The fMRI time series data were modeled by a series of events convolved with a canonical haemodynamic response function (HRF). The presentation of the feedback screen was modeled as 0-duration events. The stimuli and responses were not modeled separately as these occurred in one prior or overlapping EPI images as feedback presentation.

In the model, feedback was further subdivided into correct vs. alternative rule and positive vs. negative feedback. These trial functions were used as covariates in a general linear model, along with a basic set of cosine functions that high-pass filtered the data, and a covariate for run effects. The least-squares parameter estimates of height of the best-fitting canonical HRF for each condition were used in pair-wise contrasts. The resulting contrast images, computed on a participant-by-participant basis, were submitted to group analyses. At the group level, contrasts between conditions were computed by performing one-tailed *t*-tests on these images, treating participants as a random effect. We further performed voxelwise ANOVAs to identify regions that showed age related differences in relation to feedback processing. We tested for linear increases (−1 0 1) and decreases (1 0 −1) in the contrasts specified below.

We applied AlphaSim (Ward, 2000) to calculate the appropriate threshold significance level and cluster size for the whole-brain analyses. A significance threshold of $p < 0.05$, corrected for multiple comparisons was calculated by performing 10.000 Monte Carlo simulations in AlphaSim resulting in an uncorrected threshold of $p < 0.001$, requiring a minimum of 24 voxels in a cluster. This threshold was used for all whole-brain analyses.

We used the Marsbar toolbox for use with SPM5 (Brett et al., 2002) to perform Region of Interest (ROI) analyses to further characterize patterns of activation. We created ROIs of the regions that were identified in the functional mask of whole-brain analyses. The masks used to generate functional ROIs was based on the general (positive vs. negative feedback) contrasts ($p < 0.001$, > 24 voxels) across all participants, which was unbiased for effects of probability rule or age. Because this statistical image spanned several distinct functional brain regions in the striatum, we used Marsbar anatomical masks for the caudate nucleus to further specify our ROIs.

For all ROI analyses, effects were considered significant at an α of 0.0125, based on Bonferroni correction for multiple comparisons, $p = 0.05/4$ ROIs (caudate, DLPFC, parietal cortex and dACC), unless reported otherwise.

7.4 Results

7.4.1 Performance

To investigate the age differences in learning performance for the different stimulus pairs we calculated the percentage of correct choices (choosing the high probability stimulus) per block of 20 trials for each participant, resulting in five blocks in total. Because the two runs in the scanner consisted of new stimulus pairs, the two runs were collapsed.

As expected, the age (8–11 years, 13–16 years, 18–22 years) × probability (AB, CD) × task block (5) ANOVA showed that participants learned to make more correct choices over time, as indicated by a main effect of task block, $F(4, 260) = 40.44$, $p < 0.001$, (See Figure 7.1B). There was a significant difference in accuracy between the two probabilities; participants were more accurate on the AB (80%–20%) trials than the CD (70%–30%) trials, $F(1, 65) = 11.58$, $p < 0.001$, .Contrary to predictions, there were no age differences in learning (age × task block interaction, $F(8, 260) = 1.38$, $p = 0.11$), no age differences in accuracy on the two pairs (age × probability interaction, $F(2, 65) = 0.941$, $p = 0.393$), and no age × probability × task block interaction ($p > 0.10$). A similar ANOVA for reaction times revealed no differences for age, probability, or task block (all p's > 0.10) (see Table 7.1).

The task block factor allowed us to obtain the point in learning where participants reached a plateau. By selecting the task phase in which there were no longer differences in learning, we could examine how feedback was processed in the context of applying the correct (choosing the stimuli with a high probability of positive feedback) or alternative rule (choosing the stimuli with a low probability of positive feedback). Follow up comparisons showed that the last 60 trials were appropriate for this purpose, as performance stabilized and participants showed probability matching behavior (Shanks et al., 2002). That is, both the AB and the CD pairs showed no effects of block (learning) on accuracy in the last three blocks, $F(2, 130) = 3.47$, $p = 0.08$ and $F(2, 130) = 1.81$, $p = 0.52$, respectively. When we reanalyzed these last 60 trials, we still found a significant effect of stimulus pair, $F(1, 65) = 16.51$, $p < 0.001$, $\eta_p^2 = 0.203$, and again no significant interactions with age (all p's > 0.3).

To summarize, the behavioral results showed that all participants learned to perform more accurately over time and they learned faster on the easier AB trials than the more difficult CD trials. Performance stabilized in the last 60 trials, at which point participants showed probability matching behavior (Shanks et al., 2002).

The fMRI analyses focused on the last 60 trials. In order to have enough trial numbers in each condition, we collapsed across probabilities in the

analyses below. Thus, we differentiated between over-learned high probabilities (A and C collapsed) and alternative low probabilities (B and D trials collapsed). These will be referred to as the correct and alternative rules. Each of these rules could result in positive and negative feedback.

Table 7.2. : Brain Regions revealed by whole brain contrasts**.**

Anatomical region	L/R	voxel volume	Z	MNI coordinates		
				x	y	z
Positive > Negative						
Striatum (ventral and dorsal)	L/R	774	7.49	-6	12	-3
Dorsolateral prefrontal cortex	L	71	4.61	-27	24	51
Superior parietal cortex	L	170	4.23	-30	-75	48
Precuneus	L/R	137	4.07	-3	-36	33
Ventral Medial PFC	L/R	26	4.03	3	54	-12
Visual Cortex	L/R	332	4.50	27	-93	-9
Negative > Positive						
Dorsal Anterior Cingulate Cortex	L/R	63	4.43	9	21	36

MNI coordinators for main effects, peak voxels reported at $p < .001$, at least 24 contiguous voxels.

7.4.2 fMRI Results Positive Versus Negative Feedback

Whole-brain comparisons across age groups

First, we identified the neural correlates of feedback processing by comparing the (positive feedback vs. negative feedback) contrast across all participants. This analysis revealed increased BOLD responses for positive feedback > negative feedback in several regions including the left and right caudate, left DLPFC and left parietal cortex (see Figure 7.2A). The opposite contrast (negative > positive feedback) resulted in increased activation in the dACC. The coordinates for these comparisons (positive feedback vs. negative feedback) are reported in Table 7.2.

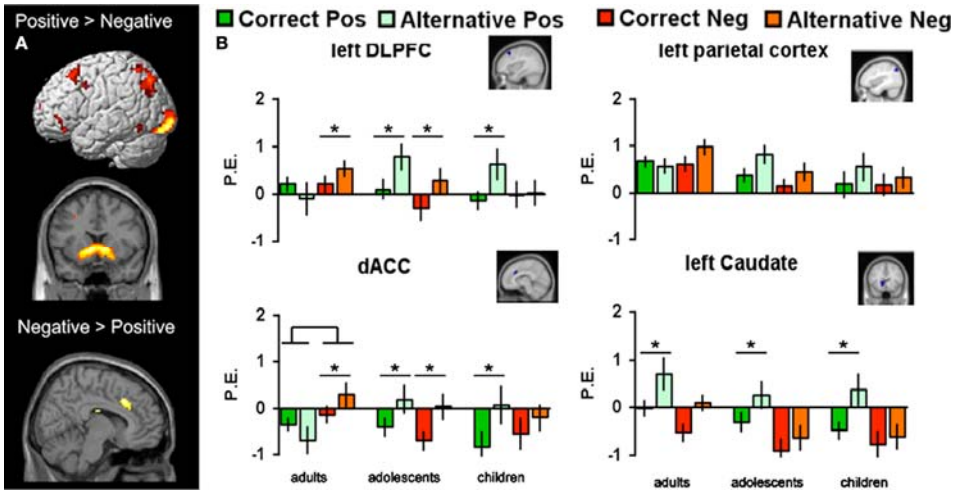


Figure 7.2 (A) Regions from the (positive vs. negative feedback) contrasts across all participants (B) Parameter estimates and standard errors for positive and negative feedback that followed either the correct or the alternative rule displayed for each age group in left DLPFC, left parietal cortex, dACC and left caudate. Significant differences between brain activity in two conditions are indicated with an asterisk (*Bonferroni corrected).

7.4.3 fMRI Region of Interest Results for Feedback × Rule × Age Group Interactions

Next, we tested for age differences and rule sensitivity in these regions by performing region of interest (ROI) analyses. The ROI analyses were restricted to the four *a priori* defined regions which emerged in the (positive vs. negative) contrast across participants: bilateral caudate, left DLPFC, left parietal cortex and dACC. In order to investigate whether there were age differences in how the statistical regularities learned by the participants had an effect on how feedback was processed we performed 3 × 2 × 2 ANOVAs testing for the interaction between valence (positive vs. negative) and rule (correct vs. alternative) as within-subjects factors and age (children, adolescents, adults) as the between-subjects factor for each ROI (see Figure 7.2B).

Left DLPFC. The (age group × valence × rule) ANOVA for left DLPFC resulted in an interaction between valence and rule, $F(2, 64) = 6.32$, $p < 0.01$, showing that left DLPFC was more active for both negative and positive feedback after choosing the alternative rule compared to the correct rule, but this difference was larger for positive than negative feedback. In addition, there was an interaction between rule (AC vs BD) and age group, $F(2, 64) = 3.87$, $p =$

0.02, and a three-way interaction between rule, valence, and age group, $F(2, 64) = 6.77$, $p < 0.01$.

As can be seen in Figure 7.2B, children and adolescents showed more activity for positive feedback after choosing the alternative rule compared to the correct rule ($t(17) = 2.64$, $p < 0.01$ and $t(26) = 3.18$, $p < 0.004$, respectively), whereas this difference was not present in adults. In addition, adults and adolescents showed more activity for negative feedback after choosing the alternative rule compared to the correct rule, ($t(21) = -2.49$, $p = 0.02$ and $t(23) = -2.81$, $p < 0.01$ respectively), but this difference was not present in children.

Left parietal cortex. The (age group × valence × rule) ANOVA for the left parietal cortex revealed a similar three-way interaction which approached significance, $F(2, 64) = 3.16$, $p = 0.05$ (see Figure 7.2B). Although the pattern of activation for the different conditions in the left parietal cortex appears similar to the pattern for left DLFPC, it did not survive Bonferroni correction and none of the *post hoc* comparisons resulted in significant effects.

dACC. The (age group × valence × rule) ANOVA for the dACC resulted in a rule × valence interaction, $F(2, 64) = 14.14$, $p < 0.001$, an age × valence interaction, $F(2, 64) = 4.11$, $p < 0.01$, and an age × rule interaction, $F(2, 64) = 4.81$, $p = 0.03$, but the three-way interaction failed to reach significance $F(2, 64) = 0.28$, $p = 0.75$.

As can be seen in Figure 7.2B, adults showed more activation in dACC after negative feedback than after positive feedback, $F(1, 21) = 8.25$, $p < 0.01$, but this was not found for the younger age groups. Children and adolescence, in contrast, showed more dACC activation after positive feedback for the alternative rule relative to the correct rule ($t(17) = 2.51$, $p < 0.01$ and $t(26) = 3.44$, $p < 0.01$ respectively). In addition, adults and adolescents showed more activity for negative feedback after choosing the alternative rule compared to the correct rule, ($t(21) = -2.89$, $p < 0.01$ and $t(26) = -3.32$, $p < 0.003$ respectively), but this difference was not present in children.

Left and right caudate. Finally, we performed an (age group × valence × rule) ANOVA for the left caudate nucleus. This analyses did not reveal any age effects, but a main effect for feedback, $F(1, 64) = 33.17$, $p < 0.001$, and a feedback × rule interaction $F(2, 64) = 17.21$, $p < 0.01$. All age groups showed more activity for the alternative (low probability) compared to the correct rule (high probability) positive feedback (all p's < 0.001), but there were no additional main or interaction effects (Figure 7.2B). Similar analyses for right

caudate yielded the same results; a main effect of feedback, $F(1, 64) = 28.16$, $p < 0.005$, and a feedback × rule interaction $F(2, 64) = 19.33$, $p < 0.01$.

7.4.4 Win Stay – Lose Shift Strategies: Behavior and Brain Analyses

Finally, to further investigate differences in feedback processing we explored developmental changes in decision-making strategies on the behavioral and neural level. In order to investigate the strategy used on the task we examined how often participants chose either the same stimulus after positive feedback (win-stay) or the other stimulus after negative feedback (lose-shift). For this set of analyses we further broke down the trials based on the subsequent choice when presented with the same stimulus pair; win-stay, win-shift, lose-stay and lose-shift. The factor 'win-stay' was computed by calculating the proportion of choice repetitions following positive feedback as a function of the total number of positive feedback events. Likewise, the factor 'lose-shift' was computed by calculating the proportion of choice shifts following negative feedback as a function of the total number of negative feedback events. Because previous analyses revealed that positive and negative feedback were processed differently dependent on rule type we analyzed the sequential effects for the correct and alternative rule separately.

Task Strategy. For correct rules, the univariate ANOVAs with age group as the between-subjects factor revealed a significant age difference in lose-shift strategies, $F(2, 64) = 4.04$, $p < 0.02$ as well as in win-stay strategies, $F(2, 64) = 4.51$, $p < 0.02$ (see Figure 7.3A). These results illustrate that adults showed more optimizing behavior than adolescents and children; they stayed more often with the correct rule after positive feedback and shifted less often after negative feedback.

For the alternative rules, the univariate ANOVAs revealed no age differences for win-stay strategies, $F(2, 64) = 0.85$, $p = 0.43$, but there was a significant age difference in lose-shift strategies, $F(2, 64) = 3.91$, $p < 0.03$. In the latter case, children showed less optimal behavior compared to the adolescents and adults; surprisingly, they stayed more often with the alternative (incorrect) rule after negative feedback.

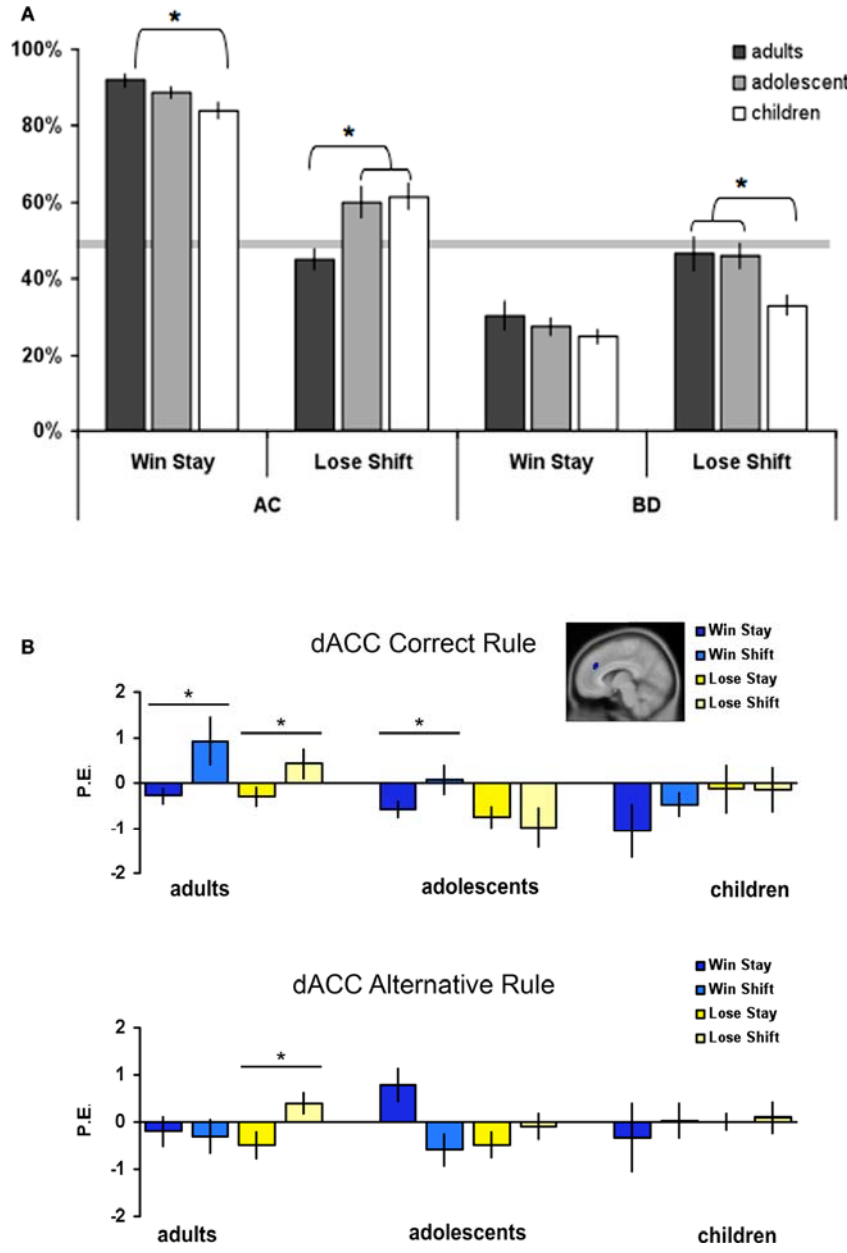


Figure 7.3:(A) Percentages of win-stay and lose-shift choices per age group and rule type, error bars represent standard error. (B) Parameter estimates and standard errors for positive and negative feedback that followed by either staying or shifting, displayed for each age group and rule type separately. Significant differences between brain activity in two conditions are indicated with an asterisk (*Bonferroni corrected).

ROI analyses. In order to explore the relation between brain activity and behavior on the subsequent trial, we compared brain activity after positive and negative feedback that resulted in staying or shifting for the two rule types separately. We explored the same ROIs as reported above. These analyses revealed significant shift and age effects only in the dACC and left DLPFC, but not in the caudate or the parietal cortex. In general, the ANOVAs showed that in adults, dACC and DLPFC were more active when participants shifted on the next trial. There were some differences in significance levels, but overall this effect seemed generally independent of feedback valence or rule. The analyses are described in more detail below.

The dACC showed the strongest relation between brain activity and subsequent behavioral change. When applying the correct rule, the shift × age group ANOVA for positive feedback revealed a main effect of shifting, $F(1, 65) = 6.27$, $p < 0.01$ but no interaction with age, $F(2, 64) = 2.29$, $p = 0.11$ (see Figure 7.3B). There was more dACC activity when shifting after positive feedback. The same ANOVA for negative feedback revealed an age × shift interaction, $F(2, 64) = 3.62$, $p = 0.03$. *Post hoc* comparisons revealed that there was more dACC activity when shifting compared to staying after negative feedback for adults ($t(21) = -2.76$, $p < 0.01$) but not for the adolescents and children (both p's > 0.1).

When applying the alternative rule, the shift × age group ANOVA for positive feedback revealed no significant effects of age or shifting. However, the same ANOVA for negative feedback revealed an age × shift interaction ($F(2, 63) = 5.31$, $p < 0.01$). *Post hoc* comparisons revealed that there was more dACC activity when shifting after negative feedback for adults ($t(21) = -3.01$, $p < 0.01$) but not for adolescents and children (both p's > 0.2).

Finally, the pattern of activation in the left DLPFC appeared similar to that of the dACC (Figure S7.2 in Supplementary Material). The shift × age ANOVAs for the correct rule resulted in significant shift × age interactions for both positive and negative feedback ($F(2, 63) = 4.46$, $p = 0.03$ and $F(2, 64) = 4.91$, $p = 0.02$, respectively). *Post hoc* test revealed that there was more left DLPFC activity when shifting on the next trial after positive and negative feedback, but this was only significant for the adults ($t(21) = -2.54$, $p < 0.01$ and $t(21) = -2.32$, $p = 0.03$, respectively). There were no significant effects for the alternative rule (all p's > 0.2).

7.4 Discussion

The goal of this study was to examine the neural developmental changes when processing positive and negative feedback signals in a probabilistic

decision-making task. As predicted, all participants learned to choose the correct rules (high probability stimuli A and C) more often than the alternative rules (low probability stimuli B and D) (Frank et al., 2004; Klein et al., 2007). After approximately 40 trials, participants adapted a performance pattern consistent with 'probability matching behavior', and this behavioral phase was the focus of our further analyses.

Behavioral analyses showed two important patterns: (1) probability matching behavior occurred in all age groups, but there were no age differences in overall learning rate, and (2) task adaptive win-stay, lose-shift strategies were observed, but age differences in adaptive behavior indicated more task-adaptive optimizing behavior in adults. These task and age differences in decision-making strategy were paralleled by changes in functional brain activity; (1) neural responses in DLPFC, dACC, and caudate were sensitive to rule × feedback interactions and an age related difference was observed in DLPFC and dACC, and (2) activity in DLPFC and dACC predicted behavioral change on subsequent trials more strongly in adults than in adolescents and children. These behavioral data and their neural correlates provide important new insights in feedback processing in general and across development. The discussion will be organized according to these themes.

7.4.1 Feedback processing in adults

Our analysis of positive and negative feedback processing in a probabilistic environment demonstrated that feedback-related activity in the DLPFC, dACC and caudate was dependent on valence and information value. We started out with a general whole-brain comparison for positive versus negative feedback and used ROI analyses to explore the areas identified in this contrast. This analysis revealed that especially left DLPFC, dACC and bilateral caudate were sensitive to feedback × rule context interactions. Before interpreting age differences in these activation patterns, we start out with the interpretation of feedback sensitivity observed in adults, which will set the stage for interpreting the developmental effects.

When exploring the data for adults separately, the results showed increased recruitment of DLPFC after receiving negative feedback following the alternative compared to the correct rule. Given that negative feedback after choosing the alternative, but not the correct, rule indicates the need for a switch in behavior, the adult findings are consistent with previous studies demonstrating negative feedback-related sensitivity in DLPFC for feedback that is important for subsequent behavioral adjustment (Kerns, 2006; van Duijvenvoorde et al., 2008; Zanolie et al., 2008) and not for negative feedback *per se*.

Besides DLPFC, the parietal cortex has previously been implicated in feedback processing (Crone et al., 2008, van Duijvenvoorde et al., 2008) and implementing cognitive control as part of the fronto-parietal network (Brass et al., 2005; Bunge et al., 2002; Dosenbach et al., 2008). In support of this hypothesis our whole-brain analyses revealed that the left superior parietal cortex was involved in feedback processing. However, in contrast with previous studies (van Duijvenvoorde et al., 2008), our subsequent *post hoc* analyses could not confirm a strong contribution of the superior parietal cortex. Possibly, the parietal cortex was more engaged in prior studies because these involved trial-to-trial learning, whereas in the current study we investigated feedback processing when rules were already learned. Future research is necessary to elucidate the role of the superior parietal cortex in feedback processing in relation to learning.

The analyses of dACC revealed a very similar activation pattern as DLPFC, however the dACC activation pattern in adults was more supportive of a general increase in activity after negative feedback regardless of rule type. Possibly, this finding indicates that, at least in adults, the dACC has a more general role in processing negative feedback; both in terms of detecting general conflict (Brown and Braver, 2005) and signaling the need for behavior change (Holroyd and Coles, 2008; Rushworth, 2008).

Finally, the caudate nucleus also showed sensitivity to feedback and rule type, but this region was more active after positive compared to negative feedback when participants chose the alternative rule. Given that this effect was specific for positive feedback, and that the probability for positive feedback for the alternative rule was low, the signal in the caudate could reflect a positive prediction error; i.e., signaling that the outcome is better than predicted (for review see Schultz, 2007).

Together, analysis of the adult activation pattern confirms prior findings showing that DLPFC and dACC are sensitive to negative feedback and the caudate is sensitive to positive feedback, but the findings further elucidate that these neural responses are dependent on the extent to which these feedback signals provide a learning signal of future performance. That is, DLPFC and caudate responses were more pronounced after selecting the incorrect rule which had a low probability of resulting in positive feedback, but which may have been important to explore. In contrast, when applying over-learned high probability rules, DLPFC and caudate were less involved, possibly because the informative value was smaller.

7.4.2 Feedback Processing: Developmental Comparisons

The neural activation patterns described above were differentially sensitive to age modulations. The first notable finding is that of differential activation patterns in the DLPFC. All participants, regardless of age, showed increased recruitment of DLPFC when choosing the alternative rule compared to the correct rule. However, children, but not adults, showed more activation in DLPFC after positive feedback when choosing the alternative rule. In contrast, adults, but not children, showed more activation in DLPFC after negative feedback when choosing the alternative rule. Adolescents seemed to be in a transition phase, because their neural response to positive feedback was similar to that observed in children, but their neural response to negative feedback was similar to that observed in adults. Thus, consistent with prior studies, these developmental differences indicate a shift from focus on positive to a focus on negative feedback with age (Somsen, 2007; Crone et al., 2008; van Duijvenvoorde et al., 2008), which appears to continue across adolescence. In addition, the current results extend previous findings by showing that developmental differences in neural responses to feedback are not related to valence *per se*, but suggest an age related change in processing learning signals with different informative value.

In contrast, for all age groups the caudate nucleus was more active for positive compared to negative feedback, in particular when participants chose the alternative rule. This finding indicates that part of the feedback processing network, which is implicated in processing statistical regularities of reward (Schultz, 2007) matures already at an early age, whereas the part of the network that is involved in processing negative feedback and the subsequent control of behavior has a more protracted developmental time course. These findings are consistent with prior reports using cognitive tasks, as these studies have also reported early maturation of subcortical regions and protracted development of cortical brain areas (Casey et al., 2004; van Duijvenvoorde et al., 2008; Velanova et al., 2008). It should be noted that other developmental studies have reported increased sensitivity of the striatum in early adolescence, however, these studies have employed paradigms with a more affective content, such as gambling tasks with real monetary rewards or emotion recognition (Ernst et al., 2005; Galvan et al., 2006; McClure-Tone et al., 2008; van Leijenhorst et al., 2009). In future studies, it will be of interest to examine whether the caudate activation can be modulated by the use of affective task modulations when learning rules or processing performance feedback.

7.4.3 Adaptive Behavior and Brain Activation across Development

One of the challenging questions for future studies is how the neural activation is associated with trial-to-trial learning. For example, we did not observe age differences in general learning performance, despite differences in neural activation. This was unexpected, and again demonstrates that differences in neural activation can be present without differences in observable behavior (Ladouceur et al., 2004). However, consistent with prior studies, the sequential analyses revealed that with age, participants became better at using the negative feedback signals to adjust their behavior on subsequent trials (Crone and van der Molen, 2004). As expected, when receiving positive feedback after having applied the correct rule, participants were more likely to stay and select the same stimulus on the subsequent trial. Likewise, when receiving negative feedback after having applied the incorrect alternative rule, participants were more likely to shift and select the correct stimulus on the subsequent trial. Overall, adults appeared better at optimizing than adolescents, and adolescents performed better than children. Based on these findings, in combination with the developmental differences in neural activation, the data are supportive of a linear increase across adolescence. Although these findings differ from earlier reports which have showed larger differences in early adolescence than in later adolescence (e.g. Ladouceur et al. 2004) the findings are consistent with prior fMRI results showing late changes in brain activation and behavior (e.g. Scherf et al., 2006; van Duijvenvoorde et al., 2008).

Intriguingly, even though children were more likely than adults to shift after receiving negative feedback when applying the correct rule, they were also more likely to stay after receiving negative feedback when applying the incorrect alternative rule. The reason for this behavioral pattern is still unclear, but it is possible that children waited with shifting when applying the incorrect alternative rule until they received positive feedback (20%). Future research should use task manipulations that allow for further investigation of this hypothesis.

We performed exploratory analyses to investigate the relation between brain activity and win-stay, lose-shift behavior, although it should be noted that these analyses are preliminary as our study design was not optimized to test for these differences. The analyses on the ROIs identified in the main analyses revealed that, consistent with prior research, dACC and left DLPFC activity predicted behavioral adjustment on the subsequent trial in adults (Kerns et al., 2004; Jocham et al., 2009). However, this pattern was observed for both rule types and appeared independent of feedback valence. Possibly, the dACC and left DLPFC were important for trial-by-trial adjustment (Kerns et al., 2004). We found a similar pattern in adolescents, but only when applying the correct rule. We

failed to find similar relations in children, which may indicate that the neural mechanisms that facilitate future behavioral adjustment are still immature or that they employed different strategies to perform the task. These interpretations are consistent with an ERP study showing increased error related negativity across adolescence (Ladouceur et al., 2007). Furthermore, the same study showed that only in adults the ERN amplitude was related to task performance.

The current study is limited by the relatively small number of trials for some of the contrasts examining the neural correlates of shifting behavior. Future studies should make use of tasks that are optimized for studying these developmental differences in more detail.

In addition, a challenging direction for future research will be to investigate the developmental differences in the learning phase. The combined use of computational reinforcement learning models (Klein et al., 2007) with imaging techniques could be a promising endeavor to parse out the developmental changes in different phases of learning (e.g. learning rate) and their neural correlates. These methods could be combined with trial-to-trial data categorization to understand how the observed developmental change in sensitivity from positive to negative feedback hinders or facilitates learning locally versus oriented towards future goals.

Conclusion

Taken together, the current findings confirm that DLPFC, dACC and caudate are important for probabilistic feedback processing, and show that they have dissociable roles as reflected in differential sensitivity to feedback valence and rule types. The DLPFC and dACC were sensitive to information value in response to negative feedback, but the caudate was sensitive to information value in response to positive feedback. These findings are consistent with previously suggested computational models of feedback learning (Cohen, 2008; Frank and Kong, 2008).

The results of this study replicate the previously reported developmental shift in sensitivity from positive to negative feedback as reflected in neural activation in the DLPFC, with a transition phase in adolescence. Using probabilistic feedback stimuli, we could dissociate between two competing hypotheses with respect to this developmental change. The results confirm the hypothesis that this shift is associated with different attention focus on learning signals and disconfirm the hypothesis that this shift reflects a simple valence effect. Further understanding of the age related changes in strategy differences, and how to influence decision-making strategies by guiding attention regulation, promise to be useful sources to improve learning behavior of children and adolescents.

7.5 Supplementary Material

7.5.1 Additional tests for Feedback x Rule and Rule x Age groups interactions

The ROI analyses presented in the manuscript suggest that neural responses to feedback valence are modulated by rule selection and age. Additionally, we performed whole-brain ANOVAs testing for interactions between valence (positive and negative), rule (correct high probability vs. alternative low probability) and age in order to explore whether additional regions were sensitive to these interactions. The whole brain ANOVA and subsequent ROI analyses of age related changes on effects of rule choice further supported the hypothesis of a shift in focus from positive feedback to negative feedback from childhood to adulthood in the DLPFC when choosing the alternative rule (see Figure S7.1).

Rule & Valence

The first ANOVA was performed to test for regions that were sensitive to the rule x valance interaction across participants. This analysis revealed a single region in the right DLPFC (BA 9, MNI: [45, 39, 30], see Figure S7.1A). This region was further explored by extracting the ROI and was found to be more active for positive feedback following the alternative rule compared to positive feedback following the correct rule, $t(58) = 4.30$, $p < .001$, and was also more active for negative feedback following the alternative rule compared to negative feedback following the correct rule, $t(63) = -3.93$, $p < .001$. In addition, a comparison of positive and negative feedback for the alternative rule indicated that the neural response was enlarged for positive feedback signals, $t(58) = 2.08$, $p < .05$.

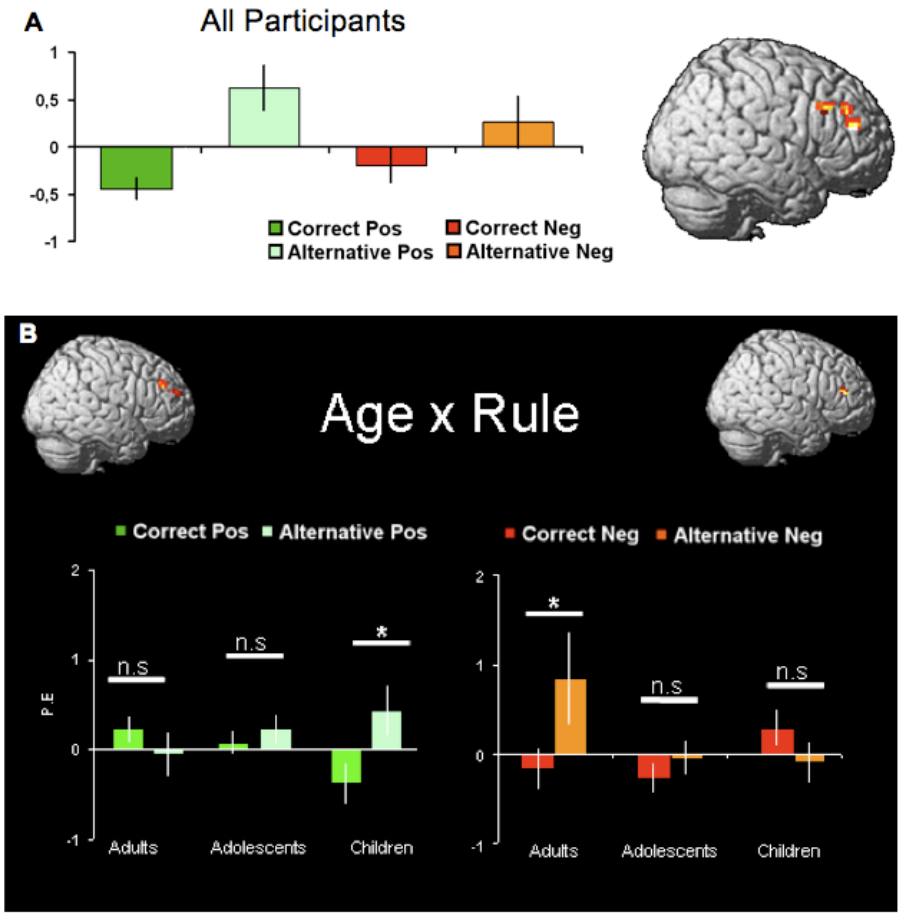


Figure S7.1: A) A whole-brain comparison revealed that the right DLPFC was sensitive to the rule x feedback interaction across all participants. Parameter estimates and standard errors for positive and negative feedback that followed either the correct or the alternative rule are displayed. (B) Regions in the right DLPFC that showed a rule x age interaction for positive and negative feedback separately. Parameter estimates and standard errors for positive and negative feedback that followed either the correct or the alternative rule are displayed for each age group. Significant differences between brain activity in two conditions are indicated with an asterisk (*).

Age x Rule for positive and negative feedback

Next, we tested for age differences by performing whole-brain ANOVAs with age group as between participants factor, testing for both linear increases [-1 0 1] as well as decreases [1 0 -1] with age. Given that differences in feedback processing were expected to differentiate based on the rule that was applied (correct vs. alternative), we tested for age differences in processing positive and

negative feedback separately. In the rule x age ANOVA for positive feedback, the decrease contrast [1 0 -1] revealed an age related change in right DLPFC (BA 9, MNI: [39, 27, 17] Figure S7.1B). The ROI of this region was extracted to test for pattern differences. Post hoc comparisons showed more activation for positive feedback after the alternative rule compared to the correct rule for children only, $t(17)= 2.93$, $p< .01$. In contrast, adults and adolescents did not show differences in processing positive feedback following the two rules (both p's >.2). No regions were detected for the increasing age contrast [-1 0 1].

The same whole-brain rule x age ANOVA was performed for negative feedback. Here, the increasing age contrast [-1 0 1] revealed a slightly lower area in the right DLPFC (BA 46, MNI: [39, 27, 17], Figure 7.3B). Subsequent post hoc comparisons for the ROI which was extracted of this region revealed increased activity for negative feedback after the alternative compared to the correct rule for adults only, $t(19)=-2.45$, $p< .01$. The adolescents and children did not show any effect of rule choice on negative feedback (both p's >.1). No regions were detected for the decreasing age contrast [1 0 -1].

In sum, the whole brain ANOVA analyses of age related changes in effects of rule choice further supported the hypothesis of a shift in focus on positive feedback to negative feedback from childhood to adulthood in the DLPFC when choosing the alternative rule. Notably, the regions which were identified in this set of ANOVAs were right lateralized. Even though the post hoc comparisons of left and right DLPFC resulted in similar activation patterns, we interpreted this difference as right DLPFC being relatively more sensitive to rule context, and left DLPFC to feedback valence.

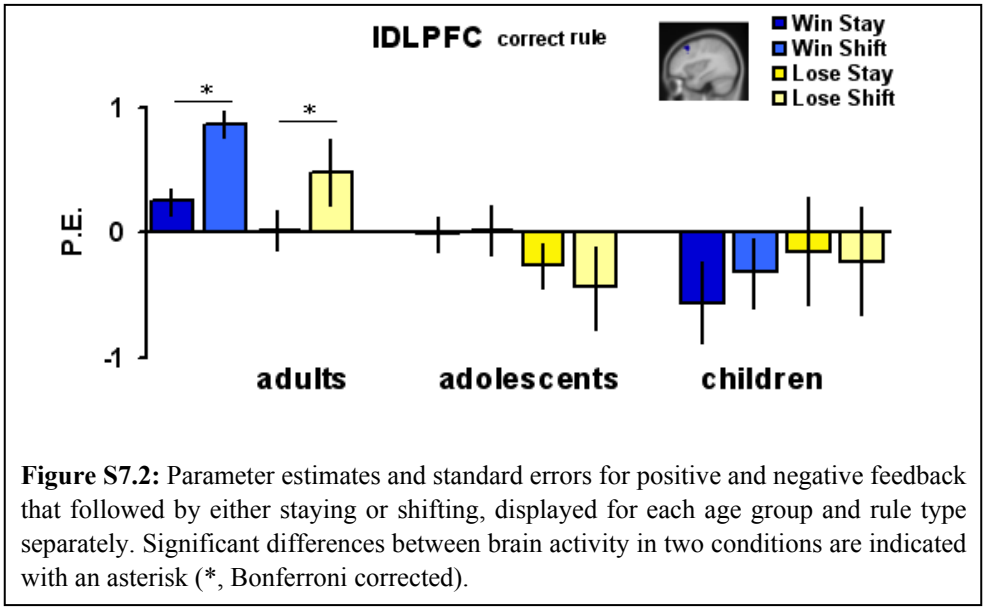


Figure S7.2: Parameter estimates and standard errors for positive and negative feedback that followed by either staying or shifting, displayed for each age group and rule type separately. Significant differences between brain activity in two conditions are indicated with an asterisk (*, Bonferroni corrected).

8. Striatum-medial prefrontal cortex connectivity predicts developmental changes in reinforcement learning

Abstract

During development, children improve in learning from feedback to adapt their behavior. However, it is still unclear which neural mechanisms might underlie these developmental changes. In the current study we used a reinforcement learning model to investigate neurodevelopmental changes in the representation and processing of learning signals. Healthy volunteers between ages 8 and 22 (children: 8–11 years, adolescents: 13–16 years, and adults: 18–22 years) performed a probabilistic learning task while in a MRI scanner. The behavioral data demonstrated age differences in learning parameters with a stronger impact of negative feedback on expected value in children. Model-based analysis of imaging data revealed that the neural representation of prediction errors was similar across age groups, but prediction error-related functional connectivity between the ventral striatum and the medial prefrontal cortex shifted as a function of age, from stronger after negative feedback to stronger after positive feedback. Furthermore, the connectivity strength predicted the tendency to alter expectations after receiving negative feedback. These findings indicate that the underlying mechanisms of developmental changes in learning may not be related to differences in the computation of learning signals per se, but rather to differences in how learning signals are used to guide behavior and expectations.

8.1 Introduction

The ability to learn contingencies between actions and positive or negative outcomes in a dynamic environment forms the foundation of adaptive behavior (Rushworth & Behrens, 2008). Learning from feedback in probabilistic environments is sensitive to developmental changes, showing developmental improvements in learning from positive and negative feedback are observed until early adulthood (Crone & van der Molen, 2004; Hooper et al., 2004; Huizinga et al., 2006; van den Bos et al., 2009). Intriguingly, prior neuroimaging studies have demonstrated developmental differences in neural circuits associated with learning from feedback in a fixed, or static learning

environment (van Duijvenvoorde et al., 2008, Crone et al., 2008). These studies show that dorsolateral prefrontal cortex and parietal cortex are increasingly engaged when receiving negative feedback. However, in a probabilistic learning environment, learning is adaptive over trials and both positive and negative feedback informs future behavior. Therefore, an important question concerns the neural mechanisms that underlie developmental differences in adaptive probability learning.

A crucial aspect of adaptive learning is using feedback to estimate the expected value of the available options. The first step in estimating the expected value is the computation of prediction errors, that is, calculating the difference between expected and experienced outcomes. Prediction errors can be positive, indicating that outcomes are better than expected, or negative, indicating that outcomes are worse than expected (Sutton & Barto, 1998). Next, these prediction errors are used to update the expected value associated with the chosen option: the expected value increases when the prediction error is positive and decreases when the prediction error is negative.

Prior neuroimaging studies have shown that activity in the ventral striatum, a target area of dopaminergic midbrain neurons, correlates with positive and negative prediction errors (Knutson et al., 2000; Pagnoni et al., 2002; e.g. McClure et al., 2003; O'Doherty et al., 2003; McClure et al., 2004). The relation between prediction errors and subsequent learning is confirmed by studies demonstrating an association between the representation of prediction errors in the striatum and individual differences in performance on probabilistic learning tasks (Pessiglione et al., 2006; Schönberg et al., 2007). Recently, a developmental study revealed heightened sensitivity in the striatum to positive prediction errors in adolescents relative to children and adults (Cohen et al., 2010). Children (ages 8-12) did not show evidence for a prediction error signal in the striatum, whereas adolescents (ages 14-19) and adults (25-30) did. Therefore, it is possible that the representation of prediction errors is one mechanism contributing to the observed developmental changes in adaptive behavior.

Several neuroimaging studies have shown that activity in the medial prefrontal cortex (mPFC) correlates with the expected value of stimuli or actions (for review see Rangel et al., 2008). Representations of expected values in the mPFC are thought to be updated by means of fronto-striatal connections, relating striatal prediction errors to medial prefrontal representations (Houk & Wise, 1995; Pasupathy & Miller, 2005; Frank & Claus, 2006; Camara et al., 2009). In support of this hypothesis, recent studies have shown increased functional connectivity between the ventral striatum and mPFC during feedback processing (Camara et al., 2008; Munte et al., 2008). Furthermore, group

differences in learning may be related to the connectivity strength between the striatum and the PFC during feedback processing. For example, substance-dependent individuals have an intact striatal representation of prediction errors, but are impaired in subsequently using these signals for learning (Park et al., 2010). This study showed that there is a positive relation between learning speed and the strength of functional connectivity between the striatum and PFC (see also Klein et al., 2007). Therefore, a second possible mechanism that may contribute to developmental changes in adaptive behavior is an increase in striatal-mPFC connectivity. Indeed, there are also still substantial changes in anatomical connectivity between subcortical structures and the prefrontal cortex during adolescence (Supekar et al., 2009; Schmithorst & Yuan, 2010).

To test these two hypotheses, a computational model of reinforcement learning model was applied to investigate developmental differences in (a) the neural representation of prediction errors, and (b) changes in fronto-striatal connectivity. Participants of three age groups (children ages 8-11, adolescents ages 13-16 and young adults ages 18-22) performed a probabilistic learning task (Frank et al., 2004) in an MRI scanner. We expect that with age, there is an improvement in learning from probabilistic feedback (Crone & van der Molen, 2004; van den Bos et al., 2009). In order to capture age related changes in learning from positive and negative feedback separately, we use a reinforcement learning model with separate learning rates for positive and negative feedback (Kahnt et al., 2009). The individually estimated trial-by-trial prediction errors generated by this reinforcement model were subsequently used to test whether developmental differences in learning reflect functional differences in the representation of prediction errors or developmental changes in the propagation of prediction errors as measured by functional fronto-striatal connectivity (Park et al., 2010).

8.2 Material and Methods

8.2.1 Participants.

Sixty-seven healthy right-handed paid volunteers ages 8-22 participated in the fMRI experiment. Age groups were based on adolescent development stage, resulting in three age groups: children (8- to 11-year-olds, n=18; 9 female), mid-adolescents (13- to 16-year-olds, n=27; 13 female) and young adults (18- to 22-year-olds, n=22; 13 female). A chi square analysis indicated that gender distribution did not differ between age groups, X^2 (2) = .79, p = .67. All participants reported normal or corrected-to-normal vision and participants or their caregivers indicated an absence of neurological or psychiatric impairments. Participants gave informed consent for the study and all

procedures were approved by the medical ethical committee of the Leiden University Medical Center.

Participants completed two subscales (similarities and block design) of either the Wechsler Adult Intelligence Scale (WAIS) or the Wechsler Intelligence Scale for Children (WISC) in order to obtain an estimate of their intelligence quotient (Wechsler, 1991, 1997). There were no significant differences in estimated IQ scores between the different age groups, F (2, 66) = 1.63, p = .20 (see Table 7.1, p. 132).

8.2.2 Task Procedure

The procedure for the probabilistic learning task (PLT, Frank et al., 2004; van den Bos et al., 2009) was as follows: The task consisted of two stimulus pairs (called AB and CD). The stimulus pairs consisted of pictures of everyday objects (e.g., a chair and a clock). Each trial started with the presentation of one of the two stimulus pairs and subsequently the participant had to choose one (e.g., A or B). Stimuli were presented randomly on the left or the right side of the screen. Participants were instructed to choose either the left or the right stimulus by pressing a button with the index or middle finger of the right hand. Responses had to be given within a 2500 ms window, which was followed by a 1000ms feedback display (see Figure 8.1 A). If no response was given within 2500 ms, the text "too slow" was presented on the screen.

Feedback was probabilistic; choosing stimulus A led to positive feedback on 80% of AB trials, whereas choosing stimulus B led to positive feedback on 20% of these trials. The CD pair procedure was similar, but probability for reward was different; choosing stimulus C led to positive feedback on 70% of CD trials, whereas choosing stimulus D led to positive feedback on 30% in these trials.

Participants were instructed to earn as many points as possible (as indicated by receiving a positive feedback signal), but were also informed that it was not possible to receive positive feedback on every trial. After the instructions and before the scanning session, the participants played 40 practice rounds on a computer in a quiet laboratory to ensure they understood the task.

In total, the task in the scanner consisted of two blocks of 100 trials each: 50 AB trials and 50 CD trials per block. The first and the second block consisted of different sets of pictures and therefore, participants had to learn a new mapping in both task blocks. The data from the last 60 trials of each block were also reported in another study using a rule-based analysis (van den Bos et al., 2009). The duration of each block was approximately 8.5 minutes. The stimuli were presented in pseudo-random order with a jittered interstimulus interval (min=1000 ms, max=6000 ms) optimized with OptSeq2 (Dale, 1999).

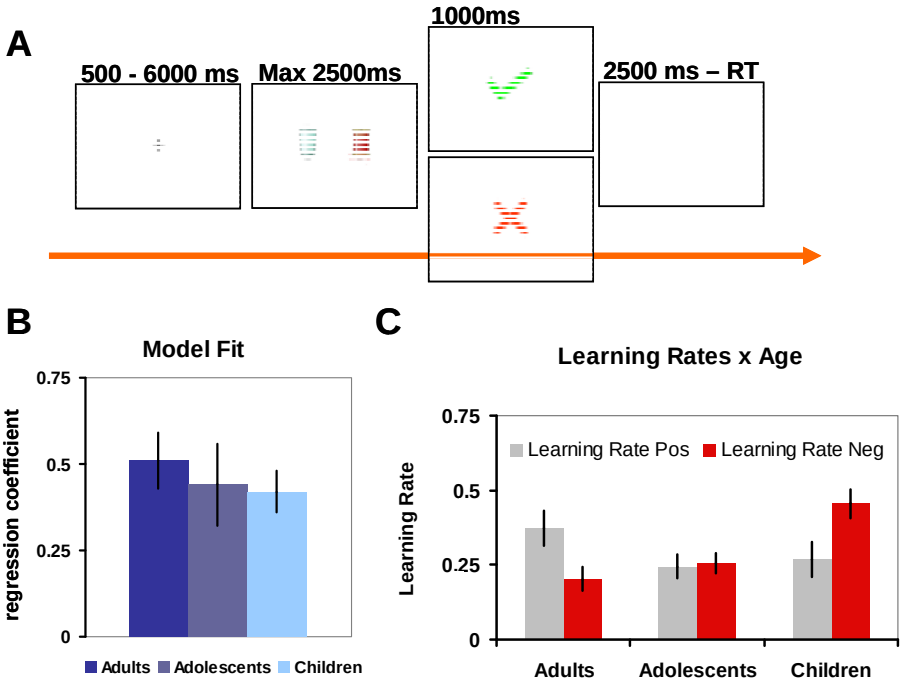


Figure 8.1: A) Participants chose one stimulus by pressing the left or right button and received positive and negative feedback according to probabilistic rules. Two pairs of stimuli were presented to the participants: (1) the AB pair with 80% positive feedback for A and 20% for B, (2) the CD pair with 70% positive feedback for A and 30% for B where A 80%-20%. **B)** Estimated learning rates for positive and negative feedback per age group. **C)** Estimated model fits per age group. Error bars represent standard error in all graphs.

8.2.3 Reinforcement Learning Model

A standard reinforcement learning model (Sutton & Barto, 1998) was used to analyze behavioral and neural data (McClure et al., 2003; Cohen & Ranganath, 2005; Haruno & Kawato, 2006; Frank & Kong, 2008; Kahnt et al., 2008). The standard reinforcement learning model uses the prediction error (δ) to update the decisions weights (*w*) associated with each stimulus (in this case A, B, C or D) (Schultz et al., 1997; Holroyd & Coles, 2002). Thus, whenever feedback is better than expected, the model will generate a positive prediction error which is used to *increase* the decision weight of the chosen stimulus (e.g. stimulus A).

However, when feedback is worse than expected, the model will generate a negative prediction error, which is used to *decrease* the decision weight of the chosen stimulus (e.g. stimulus B). The impact of the prediction error is usually scaled by the learning rate (α). We extended the standard reinforcement learning model by using separate learning rates for positive feedback (α_{pos}) and negative feedback (α_{neg}) (e.g. Kahnt et al., 2008). Thus, positive and negative feedback might have a different impact of the decisions weights. To model trial-by-trial choices, we used the soft-max mechanism to compute the probability (p) of choosing a high probability target (A or C) on trial t as the logit transform of the difference in the decision weights in each trial (w_t) associated with each stimulus, passed through a biasing sigmoid function (Montague et al., 2004; Kahnt et al., 2008). For example, when stimulus pair AB is presented the probability of choosing A is determined by:

$$p(A)_t = \frac{e^{w(A)_t}}{e^{w(A)_t} + e^{w(B)_t}} \quad (1)$$

After each decision the prediction error (δ) is calculated as the difference between the outcome received (r = 1 for positive feedback and 0 for negative feedback) and the decision weight (w_t) for the chosen stimulus:

$$\delta_t = r_t - w(chosen\_stimulus)_t \quad (2)$$

Subsequently, the decision weights are updated according to:

$$w_{t+1} = w_t + \pi \times \alpha(outcome)_t \times \delta_t \quad (3)$$

Where π is 1 for the chosen and 0 for the unchosen stimulus, α(outcome) is a set of learning rates for positive (α_{pos}) and negative feedback (α_{neg}), which scale the effect of the prediction error on the future decision weights, and thus subsequent decisions. For example, a high learning rate for positive feedback but a low learning rate for negative feedback indicates that positive feedback has a high impact on future behavior, whereas negative feedback will hardly change future behavior. These two learning rates were individually estimated by fitting the model predictions (p(high probability stimulus)) to participants' actual decisions. We used the multivariate constrained minimization function (fmincon) of the optimization toolbox implemented in MATLAB 6.5 for this

fitting procedure. Initial values for learning rates were $\alpha_{pos} = \alpha_{neg} = 0.5$ and for action values, $w(\text{left}) = w(\text{right}) = 0$.

8.2.4 Behavioral Analyses

To examine the correspondence between model predictions and participants' behavior, model predictions were compared with the actual behavior on a trial-by-trial basis. Model predictions based on estimated learning rates were regressed against the vector of participants' actual choices and individual regression coefficients were used to compare group differences in model fits. Differences in model fit between groups would indicate that other processes, for example a larger tendency to switch regardless of feedback, may play a relatively larger role in choice behavior in one group compared to the other. Only when there are no differences in model fit between groups one can confidently compare model parameters.

8.2.5 Data Acquisition

Participants were familiarized with the scanner environment on the day of the fMRI session through the use of a mock scanner, which simulated the sounds and environment of a real MRI scanner. Data were acquired using a 3.0T Philips Achieva scanner at the Leiden University Medical Center. Stimuli were projected onto a screen located at the head of the scanner bore and viewed by participants by means of a mirror mounted to the head coil assembly. First, a localizer scan was obtained for each participant. Subsequently, T2*-weighted Echo-Planar Images (EPI) (TR= 2.2 sec, TE= 30ms, 80 x 80 matrix, FOV = 220, 35 2.75mm transverse slices with 0.28mm gap) were obtained during 2 functional runs of 232 volumes each. A high-resolution T1-weighted anatomical scan and a high-resolution T2-weighted matched-bandwidth anatomical scan, with the same slice prescription as the EPIs, were obtained from each participant after the functional runs. Stimulus presentation and the timing of all stimuli and response events were acquired using E-Prime software. Head motion was restricted by using a pillow and foam inserts that surrounded the head.

8.2.6 fMRI Data Analysis

Data were preprocessed using SPM5 (Wellcome Department of Cognitive Neurology, London). The functional time series were realigned to compensate for small head movements. Translational movement parameters never exceeded 1 voxel (< 3 mm) in any direction for any subject or scan. There were no significant differences in movement parameters between age groups F (2, 65) = .15, p = .85. Functional volumes were spatially normalized to EPI templates.

The normalization algorithm used a 12 parameter affine transformation together with a nonlinear transformation involving cosine basis functions and resampled the volumes to 3 mm cubic voxels. Functional volumes were spatially smoothed using a 8 mm full-width half-maximum Gaussian kernel. The MNI305 template was used for visualization and all results are reported in the MNI305 stereotaxic space (Cosoco, Kollokian, Kwan, & Evans, 1997)

Statistical analyses were performed on individual participants' data using the general linear model in SPM5. The fMRI time series data were modeled by a series of events convolved with a canonical haemodynamic response function (HRF). The presentation of the feedback screen was modeled as 0-duration events. The stimuli and responses were not modeled separately as these occurred in one prior or overlapping EPI images as feedback presentation. To investigate the neural responses to feedback valence, independent of learning conditions, we set up a general linear model (GLM) with the onsets of each feedback type (positive and negative) as regressors.

To examine the neural correlates of reward prediction errors, we set up a second GLM with a parametric design. In this model, the stimulus functions for feedback were parametrically modulated by the trial-wise prediction errors derived from the reinforcement learning model. The modulated stick functions were again convolved with the canonical HRF. These regressors were then orthogonalized with respect to the onset regressors of positive and negative feedback trials and regressed against the BOLD signal.

Finally, to investigate age linear and quadratic age trends we applied polynomial expansion analysis (Büchel et al., 1996) with age as continuous variable, using the forward model selection as described by Büchel and colleagues (1998). Thresholds were set to $p < .001$ uncorrected for the whole group analyses, with an extend threshold of 15 continuous voxels (cf. Kahnt et al., 2008). We used the Marsbar toolbox for use with SPM5 (http://marsbar.sourceforge.net, Brett et al. 2002) to perform Region of Interest (ROI) analyses to further characterize patterns of activation and estimate individual differences in connectivity measures.

8.2.7 Functional Connectivity Analyses

To explore the interplay between the ventral striatum and other brain regions during reinforcement-guided decision-making, functional connectivity was assessed using psychophysiological interaction (PPI) analysis (Friston, 1994; Cohen et al., 2005; Cohen et al., 2008). The functional whole brain mask, in which activity correlated significantly with prediction errors for the whole group, was masked with an anatomical striatum ROI of the Marsbar toolbox that included the bilateral caudate, putamen and nucleus accumbens, to create

the seed region of interest (ROI). The method used here relies on correlations in the observed BOLD time-series data and makes no assumptions about the nature of the neural event that contributed to the BOLD signal (Cohen et al., 2008). For each model, the entire time series over the experiment was extracted from each subject in the clusters of the (left and right) ventral striatum. Regressors were then created by multiplying the normalized time series of each ROI with condition vectors that contained ones for four TRs after positive or negative prediction errors and zeros otherwise (see also Cohen & Ranganath, 2005; Kahnt et al., 2008; Park et al., 2010). Thus, the two condition vectors of positive and negative prediction errors (containing ones and zeros) were each multiplied with the time course of each ROI. These regressors were then used as covariates in subsequent analyses.

The time series between the left and right hemispheres for the ventral striatum were highly correlated (averages across runs and participants were r = .84). Therefore, parameter estimates of left- and right structures were collapsed, and thus, represent the extent to which feedback-related activity in each voxel correlates with feedback-related activity in the bilateral ventral striatum.

Individual contrast images for positive vs. negative prediction errors were computed and entered into second-level one-sample t-tests. In order to find age related differences in the whole-brain analyses of functional connectivity with the ventral striatum, we performed a regression analyses with an additional regressor for age. Thresholds for the connectivity analyses were also set to $p < .001$ uncorrected, with an extend threshold of 15 continuous voxels.

8.3 Results

8.3.1 Behavioral data

Reinforcement learning. First, we assessed how the model parameters differed between age groups. First of all, there was a good fit of the model to participants' behavior; the average regression coefficient was significantly above zero for all age groups (all p's < .001. Figure 8.1 B). Importantly, the model fit did not differ significantly between groups ($F_{(2,64)} = .96$, $p = .38$), reassuring that parameters estimations could be compared between groups.

Next, a two (learning parameters) x three (age groups) ANOVA tested for age differences in learning from positive and negative feedback. This analysis showed a significant group by parameter interaction ($F_{(2,64)} = 12.34$, $p < .001$, see Figure 8.1 C), post-hoc tests revealed that there was an age related decrease in α_{neg}, $F_{(2,67)} = 9.87$, $p < .001$, and a marginal age related increase in α_{pos}, $F_{(2,67)} = 2.96$, $p = .06$. This result indicates that the relative influence of positive

feedback on expected values decreased with age and the relative influence of negative feedback on expected values increased with age.

8.3.2 fMRI results

Model-based fMRI. Across all participants, individually generated trial-wise prediction errors (positive and negative combined) correlated with BOLD signal in bilateral ventral striatum, MPFC, posterior anterior cingulate cortex (pCC) and the bilateral amygdala extending into the parahippocampal gyrus (Figure 8.2 A, and Table 8.1). Activity in the ventral striatum was localized at an area comprising the ventral intersection between the putamen and the head of the caudate. Tests for positive and negative prediction errors separately revealed comparable results.

Whole brain regression analyses for age differences revealed no linear or non-linear age group differences (Figure 8.2 B). This analysis was repeated for positive and negative prediction errors separately and these analyses also revealed no linear or non-linear age effects. This finding shows that the prediction error (positive or negative) is not represented differently between the three age groups.

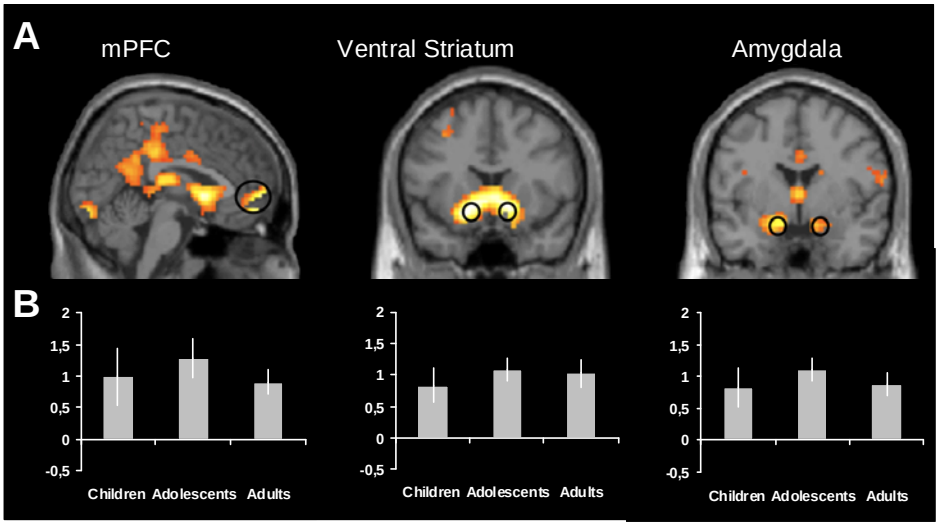


Figure 8.2: A) Regions in the medial prefrontal cortex (mPFC), ventral striatum and amygdala in which BOLD signal was significantly correlated with reward prediction errors. **B)** Parameter estimates of the prediction errors per age group in the functionally defined ROIs for the mPFC, ventral striatum and amygdala.

Table 8.1 : Brain Regions revealed by whole brain contrasts**.**

Anatomical region	L/R	BA	Z	MNI coordinates		
				x	y	z
Positive > Negative Feedback						
Ventral Striatum	L/R		7.49	-6	12	-3
Dorsolateral prefrontal cortex	L	8	4.61	-27	24	51
Superior parietal cortex	L	7	4.23	-30	-75	48
Precuneus	L/R	31	4.07	-3	-36	33
Medial PFC	L/R	10/11	4.03	3	54	-12
Visual Cortex	L/R	17	4.50	27	-93	-9
Negative > Positive Feedback						
Dorsal Anterior Cingulate Cortex	L/R	32	4.43	9	21	36
Prediction Error						
Ventral Striatum (caudate & putamen)	L/R		6.29	-6	9	3
Left Amygdala	L/R		5.50	-12	3	-18
Right Amygdala	R		5.05	18	6	-18
Medial PFC	L/R	10/11	5.84	0	54	3
Posterior Cingulate Cortex	L/R	32	4.83	0	-33	41
Visual Cortex	L/R	17	6.63	-18	-93	-18
PPI (positive > negative)						
Medial Prefrontal Cortex	L/R	10	5.47	-4	40	6
Pre-SMA	R	6	4.98	9	30	57
Right Anterior Insula / IFG	R		4.46	41	23	-9
Left Anterior Insula / IFG	L		4.67	-44	21	-3
Ventral Striatum (caudate & putamen)	L/R		7.50	9	9	3
PPI (positive > negative) x Age						
Medial PFC	L	10	4.02	-8	45	10

MNI coordinators for main effects, peak voxels reported at $p < .001$, at least 20 contiguous voxels.

Functional Connectivity. Functional connectivity between the striatum and other brain regions was assessed during processing of negative and positive feedback using PPI. The contrast used for testing functional connectivity was positive > negative feedback. Note that the vectors for positive feedback events contain all positive prediction error events, and the vectors for negative

feedback events contain all negative prediction error events. Enhanced functional connectivity was found during positive > negative feedback between the bilateral ventral striatum seed and the mPFC (Figure 8.3 A), dACC, pre-SMA, and bilateral anterior Insula extending into the inferior frontal gyrus. The opposite contrast (negative > positive feedback) did not reveal any significant changes in functional connectivity.

Next, we examined age differences in ventral striatum connectivity by adding age as a regressor to the whole-brain PPI analysis. These analyses revealed age related increases in functional connectivity of the ventral striatum seed with the mPFC (BA10) for positive > negative feedback (Figure 8.3 B). No other areas were found when testing for non-linear age effects in functional connectivity.

To further illustrate the age related changes in fronto-striatal connectivity we extracted the strength of functional connectivity between ventral striatum and mPFC for each participant and plotted it against age as a continuous variable (Figure 8.3 C). This plot reveals that the connectivity pattern shifts from a stronger connection after negative feedback for the youngest participants towards a stronger connection after positive prediction errors for the oldest participants.

Finally, we performed ROI analyses to investigate whether striatum-mPFC connectivity was related to the individual learning parameters. The differential connectivity strength (positive > negative) between the ventral striatum and mPFC ROI was used to predict the individual differences in learning rates for positive and negative feedback. The relative connectivity measure correlated negatively with the learning rate for negative feedback (r = -.39, p < .001, Figure 8.3 D), and moderately positively with the learning rate for positive feedback (r = .23, p = .07). Thus, there was stronger striatum-mPFC coupling during negative > positive feedback in participants for whom negative feedback had a relatively large impact on future expected value, whereas the reverse was true (i.e., stronger coupling during positive > negative feedback) in participants for whom positive feedback had a relatively large impact on future expected value.

To summarize, increased functional connectivity between the ventral striatum and mPFC was observed during processing of positive feedback compared to negative feedback. Furthermore, this analysis revealed that the relative strength in striatum-mPFC connectivity correlated positively with age, but negatively with the learning rate for negative feedback.

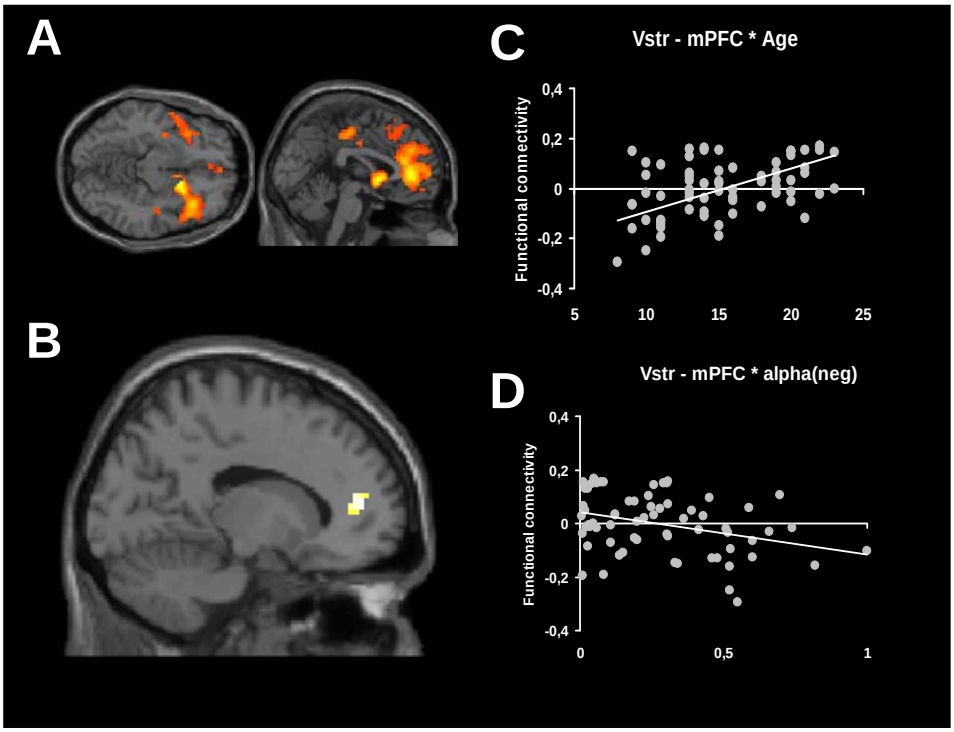


Figure 8.3: A) Regions which showed increased functional connectivity with the striatal seed region after positive compared to negative feedback. (B) Region in the mPFC that revealed age related changes in functional connectivity with the striatal seed region. Both statistical maps are all thresholded at $p < .001$, uncorrected, $k = 15$. (C) Scatterplot depicting the relationship between the functional connectivity measure of the striatum-mPFC (positive > negative feedback) and age. (D) Scatterplot depicting the relationship between the functional connectivity measure of the striatum-mPFC (positive > negative feedback) and learning rate (α_{neg}).

8.4 Discussion

The goal of this study was to examine developmental changes in the neural mechanisms of probabilistic learning. The reinforcement model showed that with increasing age, negative feedback had decreasing effects on future expected values. Imaging analyses revealed that ventral striatum activation following prediction errors did not differ between age groups; however, age differences in the learning parameters were associated with an age related increase in functional connectivity between ventral striatum and the mPFC. These behavioral data and their neural correlates allow a deeper understanding of how children, adolescents and adults learn in a changing environment. The discussion will be organized according to these themes.

Developmental changes in learning parameters

Using a reinforcement learning model we were able to disentangle differences in sensitivity to positive and negative feedback by estimating learning rates for positive and negative feedback separately. These estimated learning rates reflect the degree to which the future expected value of a stimulus will be changed after positive or negative prediction errors. As expected, the model-based analyses of learning behavior showed that with age there is a decrease in the learning rate fore negative prediction errors (α_{neg}). This finding indicates that with increasing age, the impact of negative prediction errors on the future expected value decreases. These results are consistent with developmental studies that have shown that adults are less influenced by irrelevant negative feedback (Crone et al., 2004). Furthermore, compared to younger adults, older adults have been shown to report less negative arousal to anticipated losses (Samanez-Larkin et al., 2007). Taken together, these results show that the current reinforcement model can capture the subtle age related changes in adaptive learning, and thus provides a solid basis for exploring the underlying neurodevelopment changes in representing and the processing of learning signals.

Neural Representation of prediction errors

Consistent with previous studies, trial-by-trial prediction errors generated by the reinforcement learning model correlated with activity of a network of areas including the ventral striatum, mPFC and the amygdala (Pagnoni et al., 2002; McClure et al., 2003; O'Doherty et al., 2003; Cohen & Ranganath, 2005). This result indicates that these areas are sensitive to differences in expected vs. received feedback; showing increased activation when feedback is better than expected and decreased activation when the feedback is worse than expected. Our analyses did not reveal any (linear or non-linear) age related differences in prediction errors (positive or negative). These findings are consistent with prior studies using cognitive learning tasks, which have also reported early maturation of subcortical regions and protracted development of cortical brain areas (Casey et al., 2004; van Duijvenvoorde et al., 2008; Velanova et al., 2008). However, a recent developmental study of reward-based learning using a comparable reinforcement model, with a single learning rate (for both negative and positive feedback), has shown heightened sensitivity to positive prediction errors in adolescents compared to children and adults (Cohen et al., 2010). It should be noted however, that Cohen and colleagues compared different age groups, as adolescence in this study was defined as the age range 14-19 years, and adulthood as 25-30 years. In this respect, the findings of the current study and the findings of Cohen et al. are not directly comparable. In future studies, it

will be important to test for changes in predictions errors across a wider age range and differentiating between different phases of adolescence.

The results of the current study provide different findings in comparison to affective paradigms. These studies have reported increased sensitivity of the striatum in adolescence after receiving monetary rewards or highly emotional stimuli (Galvan et al., 2006; McClure-Tone et al., 2008; Van Leijenhorst et al., 2009), which may trigger the peak in adolescent reward processing. Interestingly, Cohen et al. (2010) observed adolescent-specific increases in reaction times for 25 cents relative to 5 cents rewards. In future studies, it will be important to examine whether the prediction error representation can be modulated by the use of affective tasks or reward manipulations, and whether these effects are dependent on the development of the dopaminergic system during adolescence (for a review see Galvan, 2010).

Developmental changes in striatum-mPFC connectivity

Connectivity analyses revealed that during feedback processing the seed region in the ventral striatum sensitive to prediction errors showed increased functional connectivity with the mPFC, pre-SMA, and bilateral anterior insula/IFG during positive compared to negative feedback. This pattern of connectivity is consistent with several studies that have shown feedback-related changes in functional connectivity of the striatum (for a review see Camara et al., 2009).

Subsequent analyses revealed age related changes in striatum–mPFC functional connectivity. The pattern shifted from stronger connectivity after negative feedback for the youngest participants towards stronger connectivity after positive feedback for the oldest participants. This suggests that shifts in feedback-dependent striatum-mPFC connectivity may underlie developmental changes in learning behavior. This interpretation is in line with an adult study which has shown that the strength of ventral striatum-mPFC connectivity following feedback is related to the adjustment of behavior on subsequent trials (Camara et al., 2008). This hypothesis is further supported by the correlation between striatum-mPFC connectivity and estimated learning rate parameter for negative prediction errors in the current study.

Given that during adolescent development there are still substantial changes in structural connectivity within the prefrontal cortex (Schmithorst & Yuan, 2010) it could be hypothesized that the developmental differences in striatum-mPFC functional connectivity are related to changes in structural connectivity between these two structures (Cohen et al., 2008). In future developmental studies, it will be of interest to combine measures of structural and functional connectivity in order to further explore this hypothesis.

Additionally, it should be noted that the functional connectivity measure is uninformative about the directionality of the influence between different brain regions (Friston, 1994). Applying methods such as structural equation modeling and dynamic causal modeling (Friston et al., 1997), which take directionality into account, could further increase our knowledge of the underlying mechanisms of developmental changes in adaptive learning.

A final question concerns how these results relate to previous developmental studies of feedback processing in static environments (van Duijvenvoorde et al., 2008; Crone et al., 2008). Learning theories have suggested two separate systems that operate in parallel during feedback learning (Dickinson & Balleine, 2002); one system that operates on task explicit representations, such as rules, and another system that operates on statistical contingencies of the environment, such as feedback probabilities. Recently, a study showed that updating task representations relies on the DLPFC-parietal network, whereas updating feedback probabilities was associated with the striatum (Gläscher et al., 2010). Thus, it is likely that developmental changes in the DLPFC-parietal network represent differences in the learning system that operates on rule-based task representations, whereas the current study shows developmental differences in the system tracking statistical contingencies (see also Galvan et al., 2006; Cohen et al., 2010). The challenge for future developmental studies will be to disentangle the relative contributions of these networks, and to understand how these two systems contribute to developmental changes in feedback learning.

Conclusion

Previous studies have shown that either changes in the representation of the prediction errors in the striatum (Schönberg et al., 2007) or the connectivity of the ventral striatum with the prefrontal cortex (Klein et al., 2007; Park et al., 2010) are related to individual differences in feedback learning. In the current study we provide evidence that developmental differences in feedback learning may not be due to differences in the representation of the prediction errors per se, but rather to developmental changes in the functional connectivity between the striatum and the mPFC. This finding suggests that children do not differ in their ability to track the statistical contingencies in the task, but rather process the learning signals differently. These findings advance our understanding of the neurodevelopmental underpinnings of probabilistic learning and highlight the importance of studying neural circuits in addition to specific brain regions (Camara et al., 2009).

9. Summary and Future Directions

9.1 Introduction

The research described in this thesis concerned the development of functionally defined brain networks underlying important aspects thought to drive developmental changes in adolescent social decision-making. Developmental theories suggest that the changes in adolescent social decision-making are related to increasing capacities for: (1) perspective-taking (Eisenberg et al., 1995; Elkind, 1985), and (2) the regulation of social behavior (Steinberg, 2009). More recently it has been shown that these developmental changes in social decision-making are paralleled by substantial changes in brain structure (Giedd et al., 1999). Neurodevelopmental models hypothesize that changes in brain structure and social behavior are mediated via changes in brain function (Blakemore, 2008; Johnson, 2011).

Current neuroscientific models of interactive social decision-making suggest that there are multiple systems that contribute to social behavior; a specific 'social brain' network involved in understanding others' beliefs and intentions, and brain networks with a more general role in the monitoring and adaptation of behavior (Sanfey, 2007). Additionally, there is evidence that there are developmental changes in the activation patterns within these networks across adolescence (Blakemore, 2008; Sommerville & Casey, 2010)

The experiments in this thesis set out to test the hypothesis that the age-related changes in perspective-taking and self-regulation are associated with developmental changes in respectively the 'social brain' network, and the networks involved in the monitoring and regulation of behavior.

The first empirical study described in **Chapter 2** had two main goals: (1) to develop a new version of the Trust Game that enabled us to examine the developmental trajectory of trust and reciprocity during adolescent development, and (2) to examine the extent to which these processes are sensitive to social perspective-taking skills as measured by the risk and benefit manipulations. Participants of four age groups between 9 and 25 years participated in this study. For this study, a child friendly Trust Game paradigm was designed to capture individual and developmental differences in

perspective-taking. To examine the role of perspective-taking, experimental manipulations were added to the original Trust Game that revealed whether participants were taking the intentions of others, and consequences for others, into account (cf. Pillutla et al., 2003; Malhotra, 2004). All participants played multiple rounds of the Developmental Trust Game, in the roles of player 1 and 2, with a different anonymous other player each round. As anticipated, the results demonstrated that during development there was a general increase of both trust and reciprocity. The results of this study also demonstrated that developmental differences in trust and reciprocity depended on the extent to which the other person's perspective was taken into account. Although all age groups were more willing to trust when the risk was low rather than high, there were age related changes in sensitivity to the benefit of the other player in trust decisions; only the oldest participants were more willing to trust when the benefit for player 2 was high. Similarly, all age groups, except the youngest, were more willing to reciprocate when the benefit was high. However, only from mid adolescence onwards were participants also more willing to reciprocate when the risk for player 1 was high. The age differences in sensitivity to risk and benefit for trust and reciprocity support the hypothesis that besides a general increase of prosocial behavior, considering the outcomes for the other becomes important in social decision-making during adolescent development.

Chapter 3 describes the second empirical study with the Developmental Trust Game. The goal of this study was to investigate the neural correlates of reciprocity motives in brain regions that have previously been associated with mentalizing (aMPFC, TPJ), affective processes (ventral striatum and insula) and regulation of selfish impulses (ACC, DLPFC) in social behavior. This study was inspired by the previous findings that decisions to reciprocate trust are not only motivated by personal outcome considerations but also involve considerations of the intentions of others, and the general tendency of individuals to value the outcome of others (McClintock and Allison, 1989; de Dreu and van Lange, 1995; van Lange et al., 1997). In this study, young adults between 18 and 22 years of age were the second player in the Developmental Trust Game while fMRI data were collected.

As expected, the behavioral results showed that participants reciprocated more when the first player took a high risk to trust, indicating that participants took the consequences for the other into account. The imaging analyses revealed that two important areas of the social brain network, the aMPFC and right TPJ (Frith and Frith, 2003) have separable functions in reciprocal behavior. Consistent with previous studies, the aMPFC was more active when participants

defected compared to when they reciprocated (Gallagher et al., 2002; Decety et al., 2004). This result is consistent with the hypothesis that the aMPFC is important for self-referential processing (Northoff et al., 2006; Ochsner, 2008). In contrast to the aMPFC, the right TPJ was not sensitive to the type of choice but was sensitive to the risk manipulation when reciprocating. This result indicates that the right TPJ is involved in the shifting attention from the self to the other (Lamm et al., 2007), i.e. perspective-taking.

Further analyses showed that the ACC and the right DLPFC were most active when social impulse control was required; both these areas were activated when participants reciprocated even though the benefit of being trusted was low. In other words, when the external incentive to reciprocate was low, the ACC and the right DLPFC were more engaged in reciprocal decisions.

Finally, further analyses demonstrated that activity in the insula was sensitive to individual differences in social value orientation. The insula was more active when prosocial participants defected and more active when proself participants reciprocated. Additionally, the insula showed sensitivity to the risk manipulation; it was more active on those trials where participants chose to reciprocate when the risk that the first player took was low. Taken together, these results indicate that the insula was most active when a norm was violated (which can be a reciprocate norm for prosocial individuals or a defect norm for proself individuals, Singer et al., 2006; Montague and Lohrenz, 2007).

Chapter 4 aimed at understanding the neurodevelopmental differences in the brain areas involved in reciprocal exchange and perspective-taking. To test the neural correlates of reciprocating behavior during adolescence, a neuroimaging study was performed with the Developmental Trust Game that included adolescents and adults between ages 12 and 22 years. Using the same Developmental Trust Game the developmental changes in neural correlates of perspective-taking in reciprocal behavior were investigated.

The results of this study revealed that with age, adolescents were increasingly sensitive to the perspective of the other player as indicated by their reciprocal behavior in the different risk conditions. Furthermore, these advanced forms of perspective-taking were associated with an increased involvement of the left TPJ when being trusted. In contrast, the aMPFC was more active for the youngest participants. These results are consistent with recent developmental studies that indicated that there is an age related shift in relative contribution of the aMPFC and the TPJ during theory-of-mind tasks (e.g. reading stories, thinking about others; Wang et al., 2006; Pfeifer et al., 2007; Blakemore, 2008). Additionally, these results support the hypothesis that this shift in balance from

aMPFC to TPJ is related to a decrease in self-referential thought and an increased focus of attention on the other in social decision-making.

This study also revealed that young adults, when receiving trust, showed increased activity in the right DLPFC, an area previously found to be involved in tasks requiring cognitive control (Miller & Cohen, 2001) and the control of selfish or self-oriented impulses in the context of social dilemmas (Rilling et al., 2007). More importantly, there was an age related increase in DLPFC activity that was also related to advanced forms of perspective-taking, suggesting improved regulation of social behavior with increasing age.

Finally, this study again showed that the insula was sensitive to personal norm violations. However, in contrast to the changes in the social brain network, activity in this area did not show developmental differences, indicating this network matures at an earlier age.

In the subsequent chapter (**Chapter 5**) the neuro-developmental changes in another type of social decisions were investigated; fairness considerations. This research was inspired by prior behavioral studies that demonstrated that there are important developmental changes in perspective-taking related to fairness considerations until late adolescence (Sutter, 2007). For example, in a study using the mini-Ultimatum Game the youngest participants (9 years) were more likely to reject than to accept unfair offers, even when the proposer could not have chosen otherwise. In contrast, older participants (18 years) were more likely to accept unfair offers in that situation (Güroğlu et al., 2009).

The developmental neuroimaging study using the mini-Ultimatum Game investigated the neural correlates of age differences in fairness considerations in participants between ages 10 and 20. Consistent with prior behavioral studies, participants rejected unfair proposals when the alternative for the proposer was a fair division (Güth et al., 2008). This behavior has previously been reported in children and adults, and shows that inequity aversion motivates fairness judgment already in late childhood and early adolescence (Fehr et al., 2008; Güroğlu et al., 2009). However, children demonstrated high rejection rates for unfair offers even when the proposer did not have a fair alternative, and this rejection rate gradually dropped over the course of adolescence. These results indicate that there was an increasingly important role for taking the perspective of the other person in fairness judgments. Furthermore, the imaging analyses revealed that TPJ activity was associated with intention considerations, and that there was an age related increase in TPJ activation. Additionally, besides the TPJ, the DLPFC was also more active in adults than in children, when considering unintentional unfair offers. Finally, participants of all ages showed activation in the bilateral insula related to norm violations.

In sum, consistent with the results of the study with the Developmental Trust Game, these findings provide evidence for an early developing affective network involved in detecting norm-violations and a gradually increasing involvement of temporal and prefrontal brain regions related to intentionality considerations and the regulation of social behavior.

The study described in **Chapter 6** had two main goals: (1) to examine the development of trust relationships between late childhood and young adulthood, and (2) to examine the developmental trajectory of emotions evoked by non-cooperative behavior of others, and to what extent these emotions may lead to altruistic punishment. To investigate developmental changes in adaptive social behavior we used a repeated Trust Game paradigm in which participants, between 11 and 25 years old, interacted with three different players for several rounds (King-Casas et al., 2005). Unbeknownst to the participant the other players were computer players, preprogrammed to display different levels of trustworthiness (low, medium and high). During the repeated interactions the participants were in the role of the first player, thus, each round they had to decide whether or not to trust the other.

The data showed that adult participants often chose to trust in the first round, indicating that they expected others to reciprocate (e.g. Berg et al., 1995). In contrast, children showed a lower level of initial trust; most of them started with not trusting the other. However, for all age groups the strategy of the other player influenced the percentage of trust choices; over time all participants learned who to trust and who to distrust. Interestingly, our analyses also revealed developmental changes in strategies and adaptive behavior; all participants played a tit-for-tat type of strategy, but the children used the strictest form of tit-for-tat strategy compared to the other age groups. Further analyses revealed that children differed from adults and adolescents especially in showing higher levels negative reciprocity, thus being more sensitive to violations of trust.

Next, we investigated the relation between trust violations and participants' emotional reactions and their level of punishment. As expected, the different levels of trustworthiness displayed by the other players evoked different levels of both anger and punishment. Participants of all age groups were most angry at the player that violated trust the most and punished accordingly. Additionally, the results showed that with increasing age the amount of both anger and punishment decreased, and that age differences in trust were fully mediated by feelings of anger. Together these results indicate that the stability of adult trust relationships might be the result of an age related increase in regulation of negative affect towards violations of trust.

The studies in chapters 7 and 8 were inspired by (1) recent neuroimaging studies of social interactions that have shown that brain areas that are involved in performance monitoring are also involved in tracking and predicting the social behavior of self and other players in multi-round Games (Delgado et al., 2005; King-Casas et al., 2005; Behrens et al., 2009), and (2) developmental studies showed that monitoring and regulating behavior based on feedback signals undergoes pronounced developmental improvements between late childhood and early adulthood (Crone & van der Molen, 2004; Hooper et al., 2004). Therefore, further understanding of the age related changes in the neural mechanisms of adaptive behavior is useful for understanding developmental changes in the fundamental systems that are shown to support adaptive social behavior in multiple interactions.

In **Chapter 7** we used functional magnetic resonance imaging (fMRI) to examine the neural developmental changes when processing positive and negative feedback signals in a probabilistic decision-making task. This study was inspired by several previous studies that suggested that the neural mechanism underlying adaptive learning based on feedback signals undergo developmental changes until early adulthood (Crone et al., 2008; van Duivenvoorde et al., 2008). The study was specifically set up to test whether this developmental difference is related to valence or informative value of the feedback by examining neural responses to negative and positive feedback while applying probabilistic rules. Healthy volunteers between ages 8 and 22 years old participated in the study.

Behavioral analyses revealed that all participants learned to choose the correct rules (high probability stimuli A&C) more often than the alternative rules (low probability stimuli B&D) (Frank et al., 2004; Klein et al., 2007). After approximately 40 trials, participants adapted a performance pattern consistent with 'probability matching behavior', and this behavioral phase, consisting of the last 60 trials, was the focus of the first set of analyses. Although probability matching behavior occurred in all age groups and there were no age differences in overall accuracy, there were age differences in win-stay, lose-shift strategies. Sequential analyses revealed that the children applied a less optimal shifting strategy after negative feedback.

These age differences in decision-making strategy were paralleled by changes in functional brain activity. All participants, regardless of age, showed increased recruitment of DLPFC when choosing the alternative rule compared to the correct rule. However, children, but not adults, showed more activation in DLPFC after positive feedback when choosing the alternative rule. In contrast,

adults, but not children, showed more activation in DLPFC after negative feedback when choosing the alternative rule. Thus, consistent with prior studies, these developmental differences indicate a shift from focus on positive to a focus on negative feedback with age (Crone et al., 2008; van Duivenvoorde et al., 2008; Somsen, 2007). Taken together, these findings suggest that developmental differences in neural responses to feedback in the DLPFC are not related to valence per se, but that there is an age related change in processing learning signals with different informative value.

Chapter 8 describes a follow up study that concerned the neural mechanisms that underlie developmental differences in adaptive probability learning. In this study, based on the same data and participants as Chapter 7, we used a reinforcement learning model to investigate neurodevelopmental changes in the representation and processing of learning signals during the complete task. In order to capture age related changes in learning from positive and negative feedback separately, we use a reinforcement learning model (Sutton & Barto, 1999) with separate learning rates for positive and negative feedback (Kahnt et al., 2009). The individually estimated trial-by-trial prediction errors generated by this reinforcement model were subsequently used to test whether developmental differences in learning reflect functional differences in the representation of prediction errors or developmental changes in the propagation of prediction errors as measured by functional fronto-striatal connectivity (Park et al., 2010).

The model-based analyses of learning behavior showed that, with age, there is a decrease in the learning rate for negative feedback. This finding indicates that with increasing age, the impact of negative feedback on the future expected value decreases. Subsequent analyses of imaging data revealed that, consistent with previous studies, trial-by-trial prediction errors generated by the reinforcement learning model correlated with activity in a network of areas including the ventral striatum, mPFC and the amygdala (Pagnoni et al., 2002; McClure et al., 2003; O'Doherty et al., 2003; Cohen & Ranganath, 2005). The analyses did not reveal any age related differences in prediction errors. In contrast, age related differences in feedback adjustment were associated with increased ventral striatum connectivity with the VMPFC. The pattern shifted from stronger connectivity after negative feedback for the youngest participants towards stronger connectivity after positive feedback for the oldest participants. These findings suggest that developmental changes in adaptive behavior are not due to differences in the computation of the learning signal, but rather related to changes in how the learning signal is subsequently used in adaptive behavior.

9.2 Conclusions and Future Directions

How can these results contribute to our understanding of the relation between the development of prosocial behavior and functional brain development? Since the specific implications of the studies have been discussed in detail in the respective chapters the general discussion will take a broader perspective, focusing on theoretical and methodological points that open avenues for future inquiries.

Child's play – Games as a proxy for social development

The first important finding of the studies presented here is that the two economic games, the Trust and Ultimatum Game, capture the increased capacity of perspective-taking in relation to changes in social behavior during adolescence (Güroğlu et al., 2009; van den Bos et al., 2010). Additionally, the study employing the iterative Trust Game revealed that children use a stricter tit-for-tat strategy compared to the other age groups, especially showing increased levels of anger and retribution following trust violations. These results support the hypothesis that developmental differences in social decision-making are related to differences in capacity to regulate social feedback.

Second, the studies also yielded novel insights in the development of social behavior. As Eisenberg has shown in an extensive meta-analysis (1987), there was only a mildly positive correlation between age and prosocial behavior. Hence, many studies did not find this relationship. This raises the question to what extent age related changes in display of prosocial behavior are context-dependent. The results of the collection of studies presented in this thesis, show that economic games can be useful to further investigate this question. For instance, the study with the Developmental Trust Game suggests that from mid-adolescence onwards there is no general increase in prosocial behavior but rather a 'sophistication' of prosocial behavior. Although trust and reciprocal behavior were at a stable level at mid-adolescence, there were still changes in the effect of the outcome manipulations until late adolescence. Thus, with age, prosocial behavior becomes more context dependent, leading to more prosocial behavior in one situation but less in another. Similarly, the analyses of multiple interactions showed that children and adults showed similar responses when trust was reciprocated, but that children were more sensitive to violations of trust. These are examples of how economic games can reveal how the differences in social behavior across development are dependent on the context.

Taken together, economic games are useful extensions of the researchers' toolbox for experimental research on the development of social behavior. In future studies, economic games can further contribute to structured investigation of prosocial behavior of children, adolescents, and adults.

Neurocognitive development

The imaging studies demonstrated asynchronous developmental patterns in the 'social brain' network. In general, the pattern demonstrated a faster maturation of the aMPFC but late maturation of the TPJ. Additionally, the results showed increased involvement of the regulatory network (e.g. DLPFC), and an early maturation of the network involved in monitoring norm violations (e.g. insula). Importantly, these changes were related to developmental changes in behavior as assessed by the various social decision-making tasks. As such, the results support the hypothesis that social development is related to developmental changes in different brain networks, especially those underlying perspective-taking and self-regulation. These findings provide further support for the theoretical perspective that poses that social development is driven by increased capacities for perspective-taking and self-regulation. The following sections will: (1) reflect on the possible nature of the changes in the respective networks in light of theoretical perspectives and frameworks of brain development, and (2) point out two general directions that can advance our understanding of developmental changes in brain function.

Changing brains, changing perspectives

The analyses of the 'social brain' network identified two different developmental patterns for the aMPFC and TPJ. The aMPFC shows a pattern of local specialization, that is, in early adolescence this area is engaged in both reciprocal and defect choices, whereas from mid adolescence onwards it is only engaged in defect choices. The pattern of activity of the TPJ in both the Trust and Ultimatum Game suggests that this area gradually becomes more involved in the decision process until young adulthood. Therefore the increase in prosocial behavior might be the result of two separate processes, an early decrease in self-focus and a gradual increase in other-focus.

However, the framework of interactive specialization proposes that the developmental shift from aMPFC to TPJ may be the result of the strengthening of connections between these areas (Johnson et al., 2009). Because at younger ages the network is not fully developed young adolescents might rely more on self-reflective processes associated with the aMPFC. Findings by Blakemore and colleagues support this hypothesis; in a series of studies they showed that during adolescent development there was a developmental shift from aMPFC to TPJ activation, and at the same time an increase in connectivity strength between the aMPFC and the TPJ (Burnett et al., 2008; Burnett & Blakemore, 2009). These studies involved a passive perspective-taking task: it therefore remains to be determined whether this change in connectivity is related to the

developmental changes in social behavior. Future studies using behavioral paradigms, or re-analyses of current data-sets, are needed to investigate the role of connectivity in order to further address the nature of functional brain changes underlying social decision-making.

The regulation of social behavior

The social interaction paradigms also indicated developmental changes in the regulatory network, the DLPFC in particular. The study with the Developmental Trust Game showed that with increasing age the DLPFC gradually becomes more engaged in the decision process, showing significant relations with behavioral measures from mid-adolescence onwards. Furthermore, the data from both social interactions studies indicate that the DLPFC is engaged in situations when participants violate personal norms or behavioral tendencies. Taken together, these results fit with the theoretical accounts that the increased capacity for self-regulation is particularly driven by the gradual increase in strength of the regulatory processes to adapt social behavior (Steinberg, 2009).

The second part of this thesis had a more detailed focus on the development of the networks that underlie the monitoring and regulation of behavior in a probabilistic learning task. This section will reflect on how these results support earlier conclusions on the role of regulation in social development, but also expand on them in various ways. Finally, new hypothesis on the development of self-regulation in context of social behavior will be generated.

The initial analyses showed that the DLPFC is already involved at a young age when processing feedback in context of applying probabilistic rules. However, there was a qualitative shift in the pattern of activation, which may reflect age related changes in strategy differences and attention regulation. On the other hand, analyses of the relation between activity in the regulatory network and shifting behavior showed a very similar pattern as in the social interaction studies: there was an age related increase in the correlation between activity and behavior until young adulthood. Thus, the pattern that emerges from these data is that the DLPFC is already engaged at a young age in processing feedback from the environment, while with increasing age the relation between DLPFC activity and behavioral adaptation becomes stronger.

In subsequent analyses a reinforcement learning model was used to further explore the processes involved in adaptive behavior. These analyses revealed that age related changes in connectivity strength between the striatum and the medial PFC was related to the tendency to adjust behavior following positive or negative feedback. Taken together, these results show that age related changes in adaptive behavior are the related to developmental differences in several sub-

processes involved in monitoring and regulation, which are associated with the DLPFC/parietal cortex and striatum/mPFC networks.

Interestingly, the developmental pattern of behavior in the probabilistic learning paradigm was in one aspect very a similar to the behavior in the multiple round Trust Game, namely that children were more sensitive to negative feedback than adults. Based on this similarity in behavior, and given that the DLPFC/parietal cortex and striatum/medial PFC networks have been identified to be involved in numerous adult studies with (multiple) social interactions (Delgado et al., 2005; King-Casas et al., 2005; Behrens et al., 2009), it can be hypothesized that the reported developmental changes in brain activation will also contribute to the ability to regulate social behavior.

Consequently, it follows that the increased capacity for self-regulation of social behavior is not only due to an increased capacity to adapt future behavior, but the result of developmental changes in several sub-processes involved in self-regulation. One of the most interesting directions for future developmental studies would therefore be combining a multi round Trust Game with neuroimaging, to explore this hypothesis in more detail. The results of such studies may reveal in more detail which sub-processes of self-regulation contribute to developmental changes in social behavior.

Detecting norm violations

Finally, a very robust finding in all the social interactions studies is that all participants, almost independent of age, are sensitive to violations of social norms regarding fairness and reciprocity. This was reflected in the early maturation of the pattern of activation in the bilateral anterior insula, and by behavior in the tasks (e.g rejecting unfairness and reciprocating trust). These results suggest knowledge of these social norms is already present at the start of adolescence. Indeed, in case of fairness norms there is evidence that this already present by very young children (e.g. Fehr et al., 2008). However, the behavioral study showed that the youngest participants ages 9-10 did not always behave according to the basic norm of reciprocity, for example, when it was not in their own benefit.

Overall, these results suggest that children are already aware of social norms at a young age but predominantly react to them when it is in their own benefit. This fits well with research on the development of moral reasoning (Kohlberg, 1981) and prosocial behavior (Eisenberg et al., 1995, 2005) that suggests that young children mainly refer to selfish or hedonistic reasons when thinking about social dilemmas. By showing the early maturation of norm-violation related activity, the neuroimaging results further corroborate developmental theories that suggest that moral development during adolescence

is not a process of learning and internalizing social norms (Keller & Edelstein, 1993), but rather a process of becoming more skilled in reasoning and applying these norms (Kohlberg, 1981; Eisenberg et al., 1995, 2005). In future studies it would be interesting to expand the age range to younger populations who have not yet internalized these norms, or to investigate populations that are learning novel norms (such as at a student fraternity). One possible outcome is that in the early learning phase, norms are represented in the DLPFC/parietal network that is known to be involved in rule representation (Bunge, et al., 2009).

Multiple systems: connecting the dots

The question that remains is: how do these different networks interact? How does the information that a norm is violated, and our estimation of the intentions of the other, connect to reach a decision? Here the framework of interactive specialization points us towards a way of understanding this question in terms of brain function (Johnson, 2011). Besides the connectivity strength between brain areas within a network, the interactive specialization framework also emphasizes the importance of connectivity strength between specialized networks. In case of social behavior this could be an improved coordination between the networks that represent social norms (e.g., recognizing behavior that transgresses a norm), and the networks that are involved in taking the perspective of the other (e.g. recognizing that norm-transgressing behavior is not intentional). In support of this hypothesis, a recent study with adults showed that the functional connectivity strength between areas of the 'social brain' network (TPJ) and the affective network (VMPFC) was associated with the amount of money participants were willing to donate to charity (Hare et al., 2010). This suggests that besides an internal shift in connectivity within the 'social brain' network, developmental changes in social behavior may also be the result of strengthening of the connectivity between functional networks. Although there is no direct evidence for such a developmental pattern in the studies described in this thesis, both social interaction studies report increasing co-activation of the DLPFC and the TPJ, which might indicate a stronger functional connectivity between different networks.

To improve our understanding of the development of complex social behavior it would be beneficial to develop integrative models that describe the relation between the functional networks involved in social decision-making. The challenge for these models is not just to recognize the involvement of multiple functional networks but also to understand how these interact, for instance using network analyses (e.g. Fair et al., 2008). To conclude, measuring functional connectivity both *within* and *between* areas or networks can advance

our understanding of how these different functional networks contribute to the development of social behavior.

Computational models of social decision-making

Another promising methodological development that may contribute to our understanding of the relation between the development of cognitive processes and brain function is the use computational models (Frank et al., 2009; Poldrack, 2010). Current experimental designs allow only a limited view on the computational processes that underlie individual differences or developmental changes in behavior (Huizinga et al., 2006; Corrado & Doya, 2007). Over the past decade computational models of reward-based decision-making in combination with neuroimaging techniques have proven successful at identifying computational sub-processes and their neural implementations (for review see Rushworth & Behrens, 2008). The study in chapter 8 showed that these relatively simple models could also advance the understanding of the development of the neural mechanisms underlying monitoring and regulation of behavior based on feedback.

Recently, several studies have successfully extended these models to include processes involved in social interactions, such as predicting the mental states of others (Chang et al., 2010; Behrens et al., 2008; Hampton et al., 2008). Using these models the experimenters were able to correlate activity in brain regions with different model parameters, demonstrating dissociations between social and non-social functional processing. Additionally, these models can contribute to the understanding of how social values might interact with more basic computational processes in decision-making.

Taken together, this work shows that computational modeling in combination with neuroimaging can support stronger interpretations than what is possible using neuroimaging alone (Poldrack, 2010). Furthermore, in the past decade there has been a steady growth in the use of computational models to understand the development of cognitive functions (e.g., Mareschal, 2007; Munakata and McClelland, 2003). However, these models have not yet been integrated with neuroimaging studies of cognitive development. Future developmental studies could benefit from using computational models to gain more detailed insight in the processes that underlie changes in social behavior.

Quo vadis?

The previous part focused on (1) how the current results speak to the previous theoretical perspectives on the relation between social and brain development, and (2) how (methodologically) advancing these studies may contribute to a better understanding of the nature of social development. However, these

studies also laid groundwork for asking more challenging new questions. The next section will sketch several of those future directions in relation to the impact of internal and external influences on the development of social behavior, and how these studies can be better embedded in theoretical perspectives on social development.

Genetic and environmental influences on social behavior

Besides the developmental differences in the behavior regarding social norms, the results described in this thesis have also shown that there are large individual differences in social value orientation. These individual differences were reflected, for example, in insula activation and were similar for all age groups. Indeed, earlier studies have shown that besides developmental changes in prosocial behavior there are individual differences in prosocial attitudes that are already present at a young age and remain fairly consistent over the course of development (Eisenberg et al., 1995). One of the long standing questions for developmental and social psychology regards the exact nature of individual and developmental differences in prosocial behavior, and to what extend these are influenced by differences in genes and social environment (Lenroot et al., 2009). Currently many studies have shown that individuals differences in both genetic variables (Rueda et al., 2005) and environment (Diamond et al., 2007) are strongly associated with cognitive functioning. However, the question that remains is how these genetic and environmental differences have an impact on brain structure and function, and subsequently individual differences in behavior. For instance, it would be very interesting to be able to point out the sources, in terms of genes or environment, of the differences in neural activation between age groups that are reported in this thesis. An exciting avenue for future developmental research would therefore be combining genetics, economic games and neuroimaging to investigate the neural components of these 'hard-wired' differences in prosocial behavior, and to what extent neural differences are related to environmental variables. Note that, ultimately understanding how internal (e.g. genetic differences) and external (e.g. social economic status) factors interact and contribute to different developmental trajectories, rather than outcomes, requires longitudinal neuroimaging studies (Paus, 2010).

Hormonal changes

An example of an internal factor influencing developmental changes in behavior, that is specific to adolescence, is the influence of pubertal hormones. Numerous human and animal studies have indicated that puberty is marked by fundamental modifications in both the hypothalamic-pituitary-gonadal (HPG)

and hypothalamic-pituitary-adrenal (HPA) axes (Romeo, 2005). These pubertal shifts in HPG and HPA function result in very different levels of gonadal and adrenal steroid hormones during puberty relative to childhood and are thought to have a significant impact of brain structure and function (Ernst et al., 2008). Interestingly, these hormonal changes have also been suggested to be a driving force of developmental changes in (appetitive) social behavior (Forbes & Dahl, 2010; Nelson et al., 2005; Spear, 2000). A comprehensive perspective on social development should therefore incorporate the effects of puberty related hormonal changes. The use of economic games can be a good starting point to systemically examine the effects of puberty on social behavior. Interesting directions for future research would be the relation between pubertal hormones and: (1) developmental changes in the interactions between different sex peers (Collins, 2003), and (2) the structural and functional development of sub-cortical structures (Ernst et al., 2008; Blakemore et al., 2010).

The structure-function relationship

Linked to the previous points is the relation between brain structure and function. Although, the studies in this thesis were inspired by the changes in brain structure that take place during adolescence, they did not directly examine this topic itself. Further exploration of this relation in developmental populations can contribute to increased understanding of how internal and external factors influence brain function by re-shaping the brain. For instance, it can help determining to what extent observed age differences in brain activation reflect hard developmental constraints (e.g., anatomical constraints on signal transmission speed within certain connections). Recently, several studies have shown that that there are still significant developmental changes in structural connectivity until young adulthood (Schmithorst & Yuan, 2010), and that there are direct relations between structural connectivity and brain function (e.g Cohen, 2009, Camara et al., 2008). The multimodal analysis of structural and functional connectivity is therefore an interesting framework for understanding the relation between structural and functional development, and how network architecture shapes and constrains the development of social behavior (Honey et al., 2007; 2009).

Ecological validity

Finally, in every day life only a very small fraction of social interactions is with anonymous others. An interesting next step will therefore be to experimentally control for the relationship between the players, for instance by making use of sociometric questionnaires to identify peer relations (see Güroğlu et al., 2008). Second, behavior and neural activity associated with social interaction games

may be more strongly related to real world behavior (Rilling & Sanfey, 2010), for instance, by using experience sampling methods (Eisenberger et al., 2007). Third, future developmental studies could benefit from combining the use of games with more traditional measures (e.g. self-reports and structured interviews) of perspective taking and moral reasoning, in order to further embed the behavior in economic games in the context of existing developmental theories.

Conclusion

To conclude, this thesis describes a set of studies that have integrated research in developmental, social, and cognitive psychology, experimental economics and neuroscience. The collection of studies presented here provides to a comprehensive and multidisciplinary perspective on the development of prosocial behavior. The application of economic games yielded novel behavioral results and provided evidence for the hypothesis that developmental changes in social behavior are related to specific changes the different neural networks underlying social decision-making.

Additionally, several directions for future research were highlighted that aim at increasing our understanding of the processes and nature of developmental changes in the brain that underlie the development of social behavior. Two promising directions which can be directly applied are: (1) network/connectivity analyses, and (2) the application of computational models. The challenge for the future will be to develop an integrative model that can accommodate evidence from anatomical, functional and psychological analyses, and may account for developmental changes and individual differences in social decision-making.

Summary in Dutch

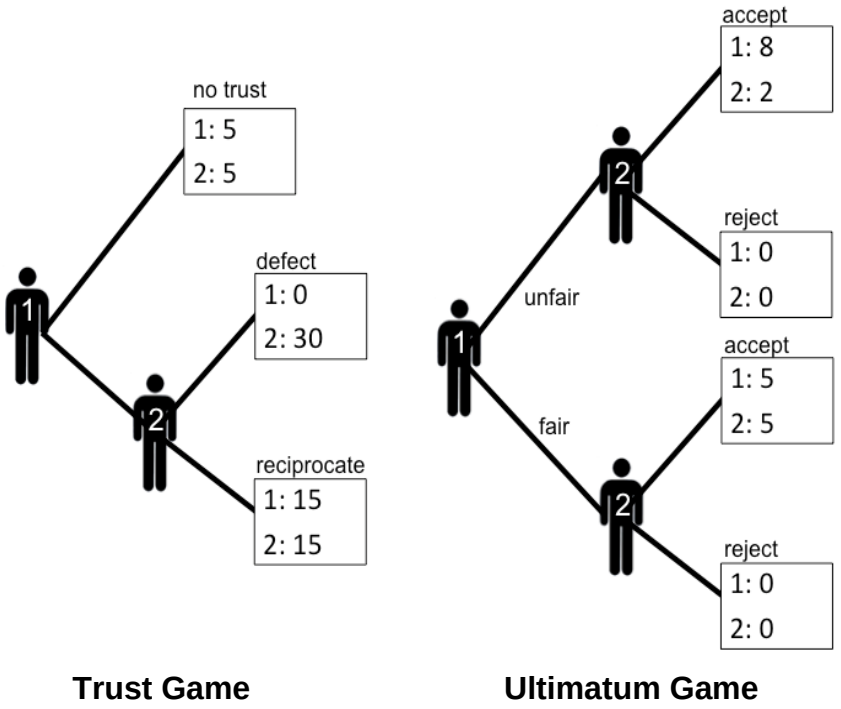
Het onderzoek beschreven in dit proefschrift is gericht op de ontwikkeling van functioneel gedefinieerde netwerken die betrokken zijn bij het maken van keuzes in een sociale context, gedurende de adolescentie. De adolescentie is een periode van grote sociale heroriëntatie; in de vroege adolescentie zijn individuen meer geneigd tot zelfgerichte gedachten en handelingen (Eisenberg et al., 1995; Elkind, 1985), terwijl zij later in de adolescentie meer geneigd zijn om aan anderen te denken, verantwoordelijkheid te nemen, en hun zelfzuchtige impulsen onder controle te houden (Steinberg, 2009). Deze veranderingen in prosociaal gedrag gaan gepaard met grote verandering in de structuur van de hersenen (Giedd et al., 1999). In het algemeen wordt door neurologische ontwikkelingsmodellen verondersteld dat de veranderingen in sociaal gedrag worden gemedieerd door veranderingen in hersenfunctie. Deze veronderstelling wordt gesteund door ontwikkelingsstudies die hebben aangetoond dat de functionele ontwikkeling *en* structurele ontwikkeling van de hersenen een zeer overeenkomstig patroon laten zien (Casey et al., 2005). Bovendien blijken deze functionele veranderingen geassocieerd te zijn met ontwikkelingsveranderingen in cognitieve functies (Crone, 2009). Echter, op dit moment is er niet veel bekend over hoe de veranderingen in de hersenen bijdragen aan specifieke veranderingen in sociaal gedrag.

Neurowetenschappelijke modellen suggereren dat er verschillende breinnetwerken zijn die bijdragen aan sociaal gedrag; het 'sociale-brein' netwerk dat betrokken is bij het begrijpen van overtuigingen en intenties van anderen, en neurale netwerken met een meer algemene rol in leren en reguleren van gedrag (Sanfey, 2007; Frank et al., 2009). Deze netwerken, die zijn geïdentificeerd in neuroimaging studies met volwassenen, functioneren als een referentiepunt voor het begrijpen van de neurologische veranderingen die ten grondslag liggen aan de ontwikkeling van sociaal gedrag. De in dit proefschrift beschreven experimenten waren gericht op het onderzoeken van ontwikkelingsveranderingen in deze specifieke functionele netwerken. Het eerste deel van dit proefschrift is gericht op de hypothese dat de ontwikkeling van sociaal gedrag gedurende de adolescentie is gerelateerd aan de toenemende

vaardigheid het perspectief van de ander in te nemen (Eisenberg et al., 1991, 1995, 2006). Daardoor ligt de focus van de eerste hoofdstukken op de ontwikkeling van het sociale-brein netwerk. Het tweede deel van dit proefschrift is gericht op de verschillen in sociale ontwikkeling in de context van herhaalde sociale interacties, en de ontwikkelingsveranderingen in de bijbehorende affectieve en regulatie netwerken. In het eerste deel van het proefschrift staan twee experimentele paradigma's van de economische speltheorie centraal; de Trust Game (het vertrouwensspel) en Ultimatum Game (het ultimatumspel). Beiden zijn zeer simpele spellen waarbij twee spelers een bepaald geld bedrag kunnen delen. In het vertrouwensspel gaat het om vertrouwen en wederkerigheid. De eerste speler in het spel krijgt een bepaald geld bedrag (10 euro) en kan kiezen om dit eerlijk te delen (5 euro voor beide spelers) of om het gehele bedrag aan de andere speler te geven (zie Figuur 9.1). Als de eerste speler alles aan de andere speler geeft dan wordt dit verdrievoudigd (het totaal is 30 euro). De tweede speler heeft nu ook weer twee keuzes. Deze kan het totale bedrag voor zichzelf houden, of dit bedrag weer eerlijk delen (15 euro voor beide spelers). De eerste speler is op de hoogte van de mogelijkheid van de tweede spelers, en zal dus alleen het geld aan de tweede speler geven al hij er op vertrouwt dat deze speler eerlijk zal gaan delen. De tweede speler is wederkerig als hij het vertrouwen van de ander beloont met het eerlijk delen van het nieuwe bedrag, en wordt als zelfzuchtig bestempeld als hij al het geld voor zichzelf houdt. Dit spel wordt vaak maar een ronde gespeeld, met anonieme spelers, de tweede speler kan dus gemakkelijk het geld voor zichzelf houden. De voorspelling vanuit economische theorie is dan ook dat mensen het geld altijd voor zichzelf zullen houden en daarom ook dat mensen als eerste speler de ander nooit zullen vertrouwen. Toch zien we dat in de werkelijkheid mensen elkaar toch vaak vertrouwen en ook dat er vaak sprake is van wederkerigheid (Berg et al., 1995).

In het ultimatumspel draait het om eerlijkheid. In dit spel zijn er ook weer twee spelers en begint de eerste speler met een bepaald geld bedrag (10 euro). De eerste speler moet de tweede een aanbod doen om dit bedrag te verdelen. De tweede speler kan dit aanbod weigeren en dan krijgen beide spelers niks, of hij kan het aanbod aannemen en dan krijgen beiden spelers uitbetaald wat de eerste speler voorstelde (zie Figuur 9.1). Economische theorie voorspelt dat mensen alle verdelingen aannemen waarbij zij meer krijgen dan 0 euro; iets is immers meer dan niets. Uit onderzoek blijkt dit niet het geval te zijn: een oneerlijke verdeling, bijvoorbeeld 8 voor mij en 2 voor jou, wordt over het algemeen geweigerd (dan liever helemaal niks!). Ook dit spel wordt over het algemeen maar een keer gespeeld dus kan er niet onderhandeld worden (de spelers kunnen elkaar ook niet zien of spreken). Hoofdstukken 2 tot en met 4 beschrijven

experimenten met de Developmental Trust Game (DTG), een kindvriendelijke versie van het vertrouwensspel (Berg et al., 1995) dat ontwikkeld is om individuele en ontwikkelingsverschillen te meten in de mate waarin proefpersonen het perspectief van de ander in acht nemen (Malhotra et al., 2004). In hoofdstuk 5 wordt een onderzoek met het mini-ultimatumspel besproken.

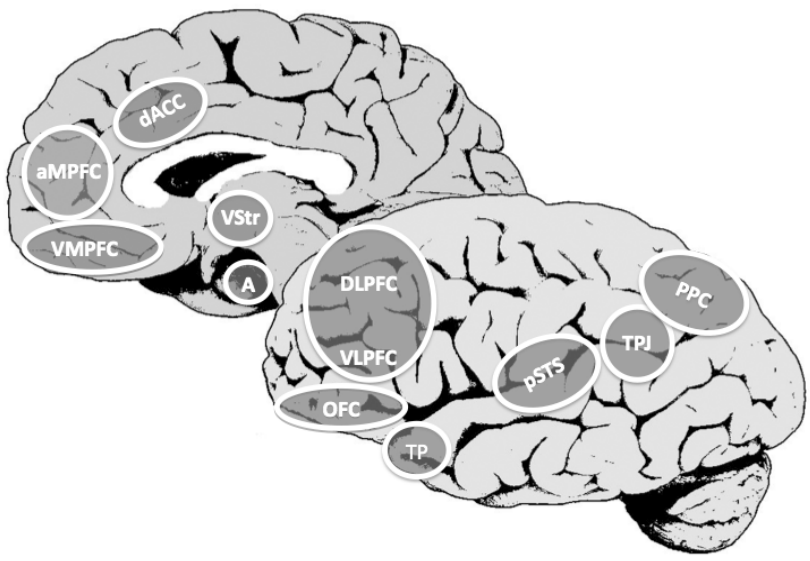


Figuur 9.1 Voorbeelden van de het vertrouwensspel en het ultimatumspel.

De studie beschreven in **hoofdstuk 2** had twee doelen: (1) om een nieuwe versie van het vertrouwensspel te ontwikkelen dat ons in staat stelde het ontwikkelingstraject van vertrouwen en wederkerigheid te onderzoeken tijdens de adolescentie, en (2) te onderzoeken in welke mate vertrouwen en wederkerigheid gevoelig zijn voor de vaardigheid het perspectief van de ander in te nemen. Vier groepen jongeren tussen 9 en 25 jaar namen deel aan deze studie. Alle deelnemers speelden meerdere rondes van het ontwikkelingsvertrouwensspel, in de rol van speler 1 en 2, telkens met een andere anonieme speler.

Zoals verwacht toonden de resultaten een algemene stijging van zowel vertrouwen als wederkerigheid tijdens de ontwikkeling. Dit resultaat geeft aan dat het ontwikkelingsvertrouwensspel in staat was om de algemene toename van

prosociaal gedrag gedurende de adolescentie, zoals beschreven in de ontwikkelingsliteratuur, vast te leggen. Daarbij hebben de resultaten van deze studie ook aangetoond dat de ontwikkelingsverschillen in vertrouwen en wederkerigheid gerelateerd waren aan de mate waarin de proefpersonen rekening hielden met het perspectief van de ander. Hoewel alle leeftijdsgroepen vaker bereid waren om de ander te vertrouwen wanneer het risico relatief klein was, waren er leeftijdsgerelateerde veranderingen in de gevoeligheid voor het voordeel van de andere speler. Alleen de oudste deelnemers waren vaker bereid om te vertrouwen wanneer het voordeel voor de andere speler relatief groot was. Alle leeftijdsgroepen, met uitzondering van de jongste, waren vaker bereid om wederkerigheid te tonen wanneer voordeel van het krijgen van vertrouwen relatief groot was. Echter, pas vanaf medio adolescentie waren deelnemers ook vaker bereid wederkerigheid te tonen als het risico voor de eerste speler relatief groot was. Deze leeftijdsverschillen in gevoeligheid voor risico's en voordelen ondersteunen de hypothese dat, naast een algemene toename van prosociaal gedrag, het perspectief van de ander steeds belangrijker wordt tijdens adolescentie.



Figuur 1.2 Schematic representation of the networks of brain areas involved in social decision-making: aMPFC = anterior Medial Prefrontal Cortex, TPJ = Temporal Parietal Junction, pSTS = posterior Superior Temporal Sulcus, TP = Temporal Poles, Vstr = Ventral Striatum, A = Amygdala, VMPFC = Ventro Medial Prefrontal Cortex OFC = Orbito frontal Cortex, dACC = dorsal Anterior Cingulate, DLPFC = Dorsolateral Prefrontal Cortrex, VLPFC = Ventrolateral Prefrontal Cortex, PPC = Posterior Parietal Cortex.

Hoofdstuk 3 beschrijft de tweede empirische studie met het ontwikkelings-vertrouwensspel. Het doel van deze studie was de neurale correlaten van individuele verschillen in wederkerigheid te onderzoeken. Deze studie was speciaal gericht op de netwerken die in verband zijn gebracht met beslissingen in sociale context; het sociale-brein netwerk (aMPFC, TPJ), het affectieve netwerk (ventrale striatum en insula) en het netwerk geassocieerd met de regulering van zelfzuchtige impulsen (ACC, DLPFC, zie figuur 9.2). Bovendien werd deze studie geïnspireerd door eerdere bevindingen dat wederkerigheid deels wordt ingegeven door individuele verschillen in de algemene tendens om de gevolgen voor anderen in acht te nemen (Sociale Waarde Oriëntatie: McClintock en Allison, 1989; De Dreu en Van Lange, 1995).

In deze studie speelden volwassen deelnemers tussen de 18 en 22 jaar de tweede speler in het ontwikkelingsvertrouwensspel terwijl zij in een MRI-scanner lagen. Zoals verwacht bleek uit onze gedragsresultaten dat de deelnemers vaker wederkerigheid vertoonden wanneer de andere speler een groot risico had genomen, wat aangeeft dat de deelnemers de gevolgen voor de andere spelers in acht namen. Uit de fMRI-analyses bleek dat de twee belangrijke gebieden van het sociale-breinnetwerk, de aMPFC en TPJ (Frith en Frith, 2003), verschillende functies hadden in wederkerig gedrag. In overeenstemming met eerdere studies was er meer activiteit in de aMPFC wanneer deelnemers voor zichzelf kozen vergeleken met wanneer zij deelden met de ander (Gallagher et al., 2002; Decety et al., 2004). Dit resultaat is in overeenstemming met de hypothese dat de aMPFC belangrijk is voor zelfgerichte processen (Northoff et al., 2006; Ochsner, 2008). In tegenstelling tot de aMPFC was de rechter TPJ niet gevoelig voor de aard van de keuze (alles houden of delen), maar wel voor de risicomanipulatie. Hieruit blijkt dat de rechter TPJ betrokken is bij het richten van aandacht op de uitkomsten voor de ander (Lamm et al., 2007).

Verdere analyses toonden aan dat de activiteit in het affectieve netwerk gevoelig was voor individuele verschillen in sociale-waardeoriëntatie. De activiteit van het striatum was hoger voor wederkerige keuzes dan voor zelfzuchtige keuzes, maar dit gold alleen voor de prosociale deelnemers. Voor de deelnemers met een zelfzuchtige waardeoriëntatie bleek juist het tegenovergestelde patroon. Deze resultaten werden geïnterpreteerd in het kader van een recente neurowetenschappelijke theorie over sociale voorkeuren. Deze theorie stelt dat voor prosociale personen wederkerigheid als een beloning wordt gezien, en dat voor zelfzuchtige individuen het materiële gewin een hogere beloningswaarde heeft. Daarbij stelt deze theorie dat het nut van sociale uitkomsten is vertegenwoordigd in het striatum (Fehr en Camerer, 2007).

Hoofdstuk 4 beschrijft ontwikkelingsverschillen in de hersengebieden die betrokken zijn bij wederkerigheid en het innemen van het perspectief van de ander. Om deze ontwikkelingsverschillen te toetsen, is er een studie gedaan met het onwikkelingsvertouwensspel met deelnemers tussen de 12 en 22 jaar. De deelnemers waren verdeeld in drie leeftijdsgroepen (12-14 jaar, 15-17 jaar en 18-22 jaar), en speelden telkens de tweede speler in het onwikkelingsvertouwensspel terwijl zij in de MRI-scanner lagen.

Uit de resultaten van deze studie bleek, net zoals in de eerdere gedragsstudie (hoofdstuk 2), dat naarmate de proefpersonen ouder werden zij gevoeliger waren voor het perspectief van de ander. Daarbij bleek ook dat deze aan leeftijd gerelateerde gevoeligheid voor het perspectief van de ander samenhing met een toename in activiteit in de linker TPJ. De activiteit in de aMPFC liet het tegenovergestelde patroon zien; deze was juist actiever voor de jongste deelnemers. Deze resultaten zijn consistent met eerdere bevindingen van ontwikkelingsstudies waaruit bleek dat kinderen en volwassenen wel hetzelfde netwerk van gebieden activeren, maar dat er een verschuiving is in activiteit binnen het netwerk van de aMPFC naar de TPJ (Wang et al., 2006; Pfeifer et al., 2007; Blakemore, 2008). Onze resultaten ondersteunen de hypothese dat deze verschuiving in de balans van aMPFC naar TPJ gerelateerd is aan een afname van zelf gericht denken en een toename in de aandacht voor de ander in sociale besluitvorming.

Deze studie toonde ook aan dat de insula gevoelig was voor het schenden van persoonlijke normen. Echter, in tegenstelling tot de veranderingen in het sociale-breinnetwerk, toont de activiteit in deze gebieden geen ontwikkelingsverschillen. Dit lijkt aan te geven dat dit netwerk al op jongere leeftijd hetzelfde functioneert als bij volwassenen. Tot slot, vonden wij ook dat voor de oudste groep DLPFC-activiteit toenam als men vertrouwd werd door de ander, terwijl dit niet het geval was bij de jongere groepen. Deze toename was tevens gerelateerd aan toenemende gevoeligheid voor het perspectief van de ander. Gezien de eerder aangetoonde rol van de DLPFC in cognitieve controle (Miller & Cohen, 2001) en de regulatie van zelfzuchtige impulsen (Riling et al., 2007), lijkt dit patroon van activiteit te wijzen op een betere regulatie van sociaal gedrag met toenemende leeftijd.

In het volgende hoofdstuk (**hoofdstuk 5**) zijn de ontwikkelingsverschillen in de hersengebieden die betrokken zijn bij het beoordelen van eerlijkheid onderzocht. Dit onderzoek was geïnspireerd op eerdere gedragsstudies die lieten zien dat kinderen al op zeer jonge leeftijd gevoelig zijn voor eerlijkheid, maar ook dat er nog belangrijke ontwikkelingen zijn in de mate dat het perspectief van de ander in deze overwegingen een rol speelt (Sutter, 2007; Guroglu et al.,

2009). Bijvoorbeeld, in een studie met het mini-ultimatumspel waren volwassen eerder geneigd om een oneerlijk aanbod te accepteren als de aanbieder hiervan geen andere keuze had, maar de jongste deelnemers (9 jaar) waren hier veel minder toe bereid.

In de neuroimaging studie met het mini-ultimatumspel hebben wij de ontwikkelingsverschillen in de hersengebieden onderzocht van deelnemers tussen de 10 en 20 jaar oud. In overeenstemming met eerdere gedragsstudies, vonden wij dat deelnemers van alle leeftijden een oneerlijk aanbod vaker afwezen wanneer er ook een eerlijk alternatief was voor de aanbieder (Guth et al., 2008). Deze resultaten ondersteunen de hypothese dat een gevoel voor eerlijkheid zich al vroeg ontwikkelt (Fehr et al., 2008; Guroglu et al., 2009.). Echter, kinderen waren, vergeleken met volwassenen, vaker geneigd om een oneerlijk aanbod af te wijzen wanneer er geen alternatief was voor de aanbieder. Dit geeft aan dat pas op latere leeftijd het perspectief van de ander ook een belangrijke rol gaat spelen in deze eerlijkheidsoverwegingen. Uit de fMRI-analyses bleek dat de activiteit in de TPJ in verband kon worden gebracht met de eerlijkheidsoverweging in het geval dat de aanbieder geen alternatieve keuze had. Daarbij bleek dat er voornamelijk verhoogde activiteit was in de TPJ wanneer dergelijke oneerlijke aanbiedingen werden afgewezen. Dit patroon van activiteit werd geïnterpreteerd als een mogelijke reflectie van schuldgevoel (Takahashi et al., 2004).

De gedrags- en imaging-resultaten samen genomen geven aan dat er (1) een leeftijdsgerelateerde toename is in de gevoeligheid voor het perspectief van de ander en (2) dat deze toename samengaat me de toenemende rol van de TPJ in eerlijkheidsoverwegingen. Bovendien waren er nog twee resultaten die zeer overeenkomstig waren met de resultaten van Hoofdstuk 4: (1) dat de DLPFC een toenemende rol kreeg in de eerlijkheidsoverwegingen naarmate de deelnemers ouder werden en (2) dat de voor alle leeftijden de activiteit in insula gerelateerd was aan het overtreden van een sociale norm. In overeenstemming met de resultaten van het ontwikkelingsvertrouwensspel leveren deze bevindingen het bewijs voor; (1) een vroegtijdige ontwikkeling van het affectieve netwerk dat betrokken is bij de opsporing van normovertredingen, en (2) een geleidelijke toename in de betrokkenheid van gebieden die gerelateerd zijn aan het innemen van het perspectief van de ander (TPJ), en de regulering van sociaal gedrag (DLPFC).

Het tweede deel van het proefschrift (Hoofdstukken 6 tot en met 8) is gericht op ontwikkelingsverschillen in het aanpassen van sociaal gedrag op basis van veranderingen in de omgeving. Het aanpassen van gedrag gebeurt vaak op basis van terugkoppeling vanuit de omgeving. Deze terugkoppeling kan positief zijn,

en het vertoonde gedrag bevorderen, of negatief zijn en juist het vertoonde gedrag ontmoedigen en aangeven dat aanpassing van gedrag nodig is. In dit deel van het proefschrift is onderzocht hoe deze aanpassingsmechanismen werken en ontwikkelen in een sociale context en hoe de neurale mechanismen die betrokken zijn bij aanpassing van gedrag zich ontwikkelen.

De studie beschreven in **hoofdstuk 6** had twee doelstellingen: (1) het onderzoeken van de ontwikkeling van adaptief sociaal gedrag in een vertrouwensspel met meerdere rondes en (2) het onderzoeken van de ontwikkelingsverschillen in de emoties die werden opgeroepen door negatieve sociale terugkoppeling. Om de ontwikkelingen in aanpassing van sociaal gedrag te bestuderen hebben wij een vertrouwensspel ontwikkeld waarbij de spelers meerdere rondes met dezelfde persoon spelen. In totaal waren er drie verschillende medespelers met wie de deelnemers dit spel speelden. De deelnemers waren altijd de eerste speler en hadden de keuze om de ander een geldbedrag toe te vertrouwen. Terwijl de deelnemers dachten online met drie anderen het spel te spelen, waren de andere spelers voorgeprogrammeerd en vertoonden verschillende niveaus van betrouwbaarheid (laag, gemiddeld en hoog). De deelnemers kwamen uit drie leeftijdsgroepen tussen de 11 en 25 jaar.

Uit de resultaten bleek dat volwassenen, vergeleken met kinderen, in het begin van het spel eerder geneigd waren om te beginnen de ander te vertrouwen. Met de tijd leerden alle deelnemers welke medespeler wel en welke niet te vertrouwen was. Toch waren er ook wel leeftijdsverschillen in aanpassingsgedrag; kinderen verschilden van de andere groepen doordat zij veel vaker negatieve wederkerigheid lieten zien. Dat wil zeggen dat zij gevoeliger waren voor het schaden van hun vertrouwen en daarna eerder geneigd waren de ander geen geld meer toe te vertouwen totdat deze goede wil had getoond (zelfs als uit de vele rondes daarvoor bleek dat de persoon zeer betrouwbaar was).

Vervolgens hebben wij ook gekeken naar de emotionele reacties op het verbreken van vertrouwen en hoe deze emoties (in het bijzonder boosheid) motiveren om de ander te straffen voor zijn asociale gedrag. Zoals verwacht bleek uit onze analyse dat deelnemers van alle leeftijden het boost waren op de persoon die het vertrouwen het vaakst had beschadigd en dat die persoon ook het hardst werd gestraft. Ook lieten de resultaten zien dat de mate van boosheid afnam als de leeftijd toenam en dat de leeftijdsverschillen in mate van straffen gemedieerd werden door de mate van boosheid. Deze resultaten geven aan dat de leeftijdsgerelateerde toename in stabiliteit van vertrouwensrelaties mogelijk te danken is aan een afname in gevoeligheid voor negatieve wederkerigheid, mogelijk door een toenemend vermogen om negatieve emoties te reguleren.
In **hoofdstuk 7** zijn de ontwikkelingsverschillen onderzocht in de neurale

correlaten van positieve en negatieve feedbackverwerking. Deze studie was gebaseerd op een aantal eerdere studies die ontwikkelingsveranderingen aantoonden in de neurale mechanismen die ten grondslag liggen aan adaptief leren (Crone et al., 2008; Van Duivenvoorde et al., 2008.). Deze studie was er op gericht om te testen of deze ontwikkelingsveranderingen in feedbackverwerking gerelateerd zijn aan de valentie (positief of negatief) of de informatieve waarde (gedrag veranderen of niet) van feedback. Gezonde vrijwilligers tussen de leeftijd van 8 en 22 jaar oud namen deel aan deze studie.

Voor dit onderzoek werd er een kindvriendelijke probabilistische leertaak ontwikkeld. Tijdens deze taak werden iedere keer twee paren van twee plaatjes getoond (het AB en het CD paar). De deelnemers moesten telkens een van de twee plaatjes uitkiezen. Vervolgens kregen de deelnemers positieve of negatieve feedback op hun keuze. In het begin wisten de deelnemers nog niets over de plaatjes maar gedurende het experiment leerden de deelnemers welke plaatjes de grootste kans hadden op positieve feedback (A en C, 80 en 70%) of negatieve feedback (B en D, 20 en 30%).

Uit de gedragsanalyses bleek dat alle deelnemers leerden om de juiste regel (plaatjes A en C) vaker te kiezen dan de alternatieve regel (plaatjes B en D). Na ongeveer 40 rondes werd het gedragspatroon van de deelnemers consistent. Hoewel de kinderen even snel leerden welke plaatjes de goede waren, bleken er wel leeftijdsverschillen te zitten in de keuzestrategieën. Uit de sequentiële analyses bleek dat de kinderen een minder optimale strategie toepasten na het krijgen van negatieve feedback. Uit de fMRI-analyses bleek dat deze leeftijdsgerelateerde verschillen in strategie gepaard gingen met veranderingen in hersenactiviteit.

Alle deelnemers, ongeacht leeftijd, vertoonden verhoogde activiteit in de DLPFC wanneer zij de alternatieve regel kozen. Echter, kinderen vertoonden meer activiteit in de DLPFC na positieve feedback bij het kiezen van de alternatieve regel en volwassenen vertoonden juist meer activiteit in de DLPFC na negatieve feedback. In overeenstemming met eerdere studies wijzen deze ontwikkelingsverschillen op een verschuiving van een focus op positieve naar een focus op negatieve feedback (Crone et al., 2008; Van Duivenvoorde et al., 2008; Somsen, 2007). Tevens laten deze bevindingen zien dat de ontwikkelingsverschillen in de neurale reacties op feedback geen verband houden met valentie per se, maar ook afhankelijk zijn van de informatieve waarde van de feedbacksignalen.

De studie in **hoofdstuk 8** beschrijft additionele analyses op de data van het experiment beschreven in hoofdstuk 7. Deze analyses maakten gebruik van een computationeel model en had als doel ontwikkelingsverschillen in de neurale

mechanismen die betrokken zijn bij leren nader te onderzoeken. De gedragsdata werden geanalyseerd door middel van een reinforcement learning model (Sutton & Barto, 1999) met verschillende leerparameters voor positieve en negatieve feedback (Kahnt et al., 2009). Een reinforcement learning model is een computationeel model dat ervan uitgaat dat tijdens het leren de verwachte uitkomst van een keuze telkens wordt aangepast op basis van de feedback. Deze aanpassing van de verwachting gaat middels een leersignaal; de prediction error of voorspellingsfout. Dit signaal kan klein of groot, positief of negatief zijn, naarmate de inschatting van de proefpersoon te laag of te hoog was vergeleken met de werkelijke uitkomst. De leerparameters in het model bepalen vervolgens in welke mate de voorspellingsfout wordt gebruikt om de verwachte waarde van een keuze aan te passen. Als de leerparameter groot is betekent dat, dat een persoon zijn verwachtingen telkens in grote mate aanpast wanneer deze uitkomst anders was dan verwacht. Is deze zeer klein dan zal deze persoon zijn verwachtingen en dus ook zijn gedrag niet snel veranderen op basis van de signalen uit de omgeving.

Uit de gedragsanalyses bleek dat met toenemende leeftijd een daling plaatsvindt in de leerparameter voor negatieve feedback. Deze bevinding geeft aan dat, met toenemende leeftijd, de impact van de negatieve feedback op de toekomstige verwachte waarde daalt. De individueel geschatte voorspellingsfouten en leerparameters, gegenereerd door het computationele model, zijn vervolgens gebruikt om ontwikkelingsverschillen in neurale processen nader te onderzoeken.

Uit de fMRI-analyses bleek dat, in overeenstemming met eerdere studies, de voorspellingsfouten correleerden met de activiteit in het ventrale striatum (Pagnoni et al., 2002; McClure et al., 2003; O'Doherty et al., 2003; Cohen & Ranganath, 2005). De analyses toonden ook aan dat er geen leeftijdsgerelateerde verschillen zijn in neurale representatie van de voorspellingsfouten. Daarentegen waren er wel leeftijdsgerelateerde verschillen in de functionele connectiviteit, oftewel in de synchronisatie van activiteit, tussen het striatum en de VMPFC. Het patroon liet een verschuiving zien van sterkere connectiviteit na negatieve feedback voor de jongste deelnemers tot sterkere connectiviteit na positieve feedback voor de oudste deelnemers. Deze bevindingen suggereren dat veranderingen in de ontwikkeling van adaptief gedrag niet te wijten zijn aan verschillen in de berekening van het leersignaal, maar veroorzaakt worden door verschillen in de manier waarop het leersignaal vervolgens wordt gebruikt om toekomstig gedrag aan te passen.

Conclusie

Dit proefschrift beschrijft een reeks van studies die gebaseerd zijn op onderzoek

in de ontwikkelings-, sociale en cognitieve psychologie in combinatie met onderzoek uit de experimentele economie en de neurowetenschappen. Deze collectie van de studies biedt een uitgebreid en multidisciplinair perspectief op de ontwikkeling van prosociaal gedrag. De toepassing van economische spellen leverde nieuwe gedragsresultaten, en ondersteuning voor de hypothese dat ontwikkelingsveranderingen in sociaal gedrag zijn gerelateerd aan veranderingen van de verschillende neurale netwerken. De belangrijkste bevinden worden hieronder nog een keer kort op een rijtje gezet.

Kinderspel – Spelen als methode voor onderzoek naar sociale ontwikkeling
De eerste belangrijke bevinding is dat de twee economische spellen, het vertrouwens- en het ultimatumspel, de gedragsveranderingen in sociaal gedrag tijdens de adolescentie, zoals beschreven in de literatuur goed konden repliceren (Güroğlu et al., 2009; van den Bos et al., 2010). Daarbij hebben de resultaten van de studies met de economische spellen ook nieuwe inzichten opgeleverd. Bijvoorbeeld, dat na midden-adolescentie gedrag niet per se meer sociaal wordt maar eerder meer context afhankelijk. Dit kan leiden tot meer prosociaal gedrag in de ene situatie, maar minder in de andere. Uit de analyses van het spel met meerdere interacties is gebleken dat kinderen vooral gevoeliger zijn voor schendingen van vertrouwen, maar op eenzelfde manier reageren op wederkerigheid. Dit zijn beiden voorbeelden van hoe de economische spellen kunnen onthullen hoe de ontwikkelingsverschillen in sociaal gedrag in ontwikkeling afhankelijk zijn van de context waarin zij plaatsvinden. In toekomstige studies kunnen deze spellen verder bijdragen aan gestructureerd onderzoek naar prosociaal gedrag van kinderen, adolescenten en volwassenen. Deze studies hebben laten zien dat de economische spellen nuttige uitbreidingen zijn van onderzoekers' instrumenten voor experimenteel onderzoek.

Veranderende hersenen, veranderende perspectieven
De analyses van het sociale-brein netwerk hebben twee verschillende ontwikkelingspatronen voor de aMPFC en TPJ geïdentificeerd. De aMPFC toont een patroon van lokale specialisatie, dat wil zeggen in de vroege adolescentie is dit gebied actief voor zowel wederkerige en zelfzuchtige keuzes, terwijl het vanaf midden-adolescentie allen activiteit vertoont bij zelfzuchtige keuzes. De TPJ wordt juist geleidelijk aan steeds meer betrokken bij het keuze proces, en deze ontwikkeling gaat door tot jong-volwassenheid. Deze resultaten suggereren dat de veranderingen in prosociaal gedrag het resultaat zijn van ontwikkelingen in twee afzonderlijke processen; (1) een vroege daling in zelf-focus (aMPFC) en (2) een geleidelijke toename in aandacht voor de ander (TPJ).

De regulering van sociaal gedrag
De studies in dit proefschrift hebben ook aangegeven dat er belangrijke ontwikkelingsveranderingen plaatsvinden in het regulatie netwerk, de DLPFC in het bijzonder. De studie met het vertrouwensspel toonde aan dat met toenemende leeftijd de DLPFC geleidelijk meer betrokken wordt in het besluitvormingsproces, en dat vanaf midden adolescentie er een sterke relatie is tussen DLPFC activiteit en de mate van prosociaal gedrag. Bovendien, de gegevens van beide sociale interactie studies geven aan dat de DLPFC zich bezighoudt met situaties waarin de deelnemers extra controle moeten uitoefenen. Deze resultaten ondersteunen de theoretische modellen die vooronderstellen dat de toenemende capaciteit voor zelfregulering een zeer belangrijke rol speelt in de ontwikkeling van sociaal gedrag (Steinberg, 2009).

Normovertredingen
Ten slotte, uit de sociale interactie studies bleek dat alle deelnemers, onafhankelijk van de leeftijd, gevoelig zijn voor schendingen van sociale normen ten aanzien van eerlijkheid en wederkerigheid. Dit kwam tot uiting in de vroege rijping van het patroon van activiteit in de bilaterale anterior insula, en door het gedrag in de spellen (bijv. afwijzing van onrechtvaardigheid en de hoge mate van wederkerigheid). Deze resultaten suggereren dat kennis van deze sociale normen al aanwezig is bij het begin van de adolescentie. Deze resultaten sluiten aan bij recente studies die laten zien dat het gevoel van eerlijkheid al op zeer jonge leeftijd aanwezig is (bijvoorbeeld Fehr et al., 2008). Het is interessant om te zien dat er tijdens de ontwikkeling, naast de verschillen in het sociale en regulerende netwerk, ook breinnetwerken zijn die geen veranderingen laten zien.

Tot Slot
De imaging studies hebben asynchrone ontwikkelingspatronen aangetoond in het netwerk van de 'sociale brein'. De resultaten toonden een snellere rijping van de aMPFC, maar late rijping van de TPJ. Daarnaast toonde de resultaten een grotere betrokkenheid van het regulerende netwerk (DLPFC), en een vroege rijping van het affectieve netwerk dat betrokken is bij normovertredingen. Deze studies hebben bijgedragen aan een dieper inzicht in de processen die ten grondslag liggen aan de sociale ontwikkeling tijdens de adolescentie. De uitdaging voor toekomstige studies is om een model te ontwikkelen om de resultaten van studies naar structurele en functionele hersenontwikkeling te integreren op een manier dat deze in staat is de psychosociale ontwikkeling te verklaren.

References

A

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Review of Neuroscience, 7*, 268-277.

Arnett J.J. (2000). Adolescence and Emerging Adulthood, A Cultural Approach (Prentice Hall).

Achenbach, T. M. (1991). Manual for the child behavior checklist 4-18 / and 1991 profile. Burlington: VT: University of Vermont, Department of Psychiatry.

B

Baumeister, R. F., & Vohs, K. D. (2007). Self-Regulation, Ego Depletion, and Motivation. *Social and Personality Psychology Compass, 1*, 1-14.

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., & Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron, 58*(4), 639-650.

Beate, S., & Frith, U. (1992). Deception and sabotage in autistic, retarded and normal children. *Journal of Child Psychology and Psychiatry, 33*(3), 591–660.

Bechara, A., A. R. Damasio, H. Damasio & S. W. Anderson (1994). "Insensitivity to future consequences following damage to the human prefrontal cortex." *Cognition, 50*(1-3), 7-15.

Behrens, T. E., L. T. Hunt & M. F. Rushworth (2009). "The computation of social behavior." *Science, 324*(5931), 1160-1164.

Behrens, T. E., L. T. Hunt, M. W. Woolrich & M. F. Rushworth (2008). "Associative learning of social value." *Nature, 456*(7219), 245-249.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10*(1), 122-142.

Bernath, M. S., & Feshbach, N. D. (1995). Children's trust: Theory, assessment, development, and research directions. *Applied and Preventive Psychology, 4*, 1–19.

Berns, G. S., S. Moore & C. M. Capra (2009). "Adolescent Engagement in Dangerous Behaviors Is Associated with Increased White Matter Maturity of Frontal Cortex." *PLoS One, 4*(8), e6773.

Bolle, F. (1995). Rewarding trust: An experimental study; working paper, Europa-Universität Viadrina.

Blakemore, S. J., & Choudhury, S. (2006). Development of the adolescent brain: Implications for executive function and social cognition. *Journal of Child Psychology and Psychiatry, 47*(3-4), 296-312.

Blakemore S.J. (2008). The social brain in adolescence. *Nature Reviews Neuroscience*, *9*, 267-77.

Blakemore S-J, Burnett S., Dahl R.E. (2010). The role of puberty in the developing adolescent brain. *Human Brain Mapping, 31*, 926-933.

Blair R.J.R., Cipolotti L. (2000). Impaired social response reversal. *Brain 123*, 1122-1141.

Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature, 402*(6758), 179-182.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*(3), 624-652.

Brass, M., Derrfuss, J., Forstmann, B., & von Cramon, D. Y. (2005). The role of the inferior frontal junction area in cognitive control. *Trends in Cognitive Sciences, 9*(7), 314-316.

Brett M., Anton J.L., Valabregue R., & Poline J.B. (2002) Region of interest analysis using an SPM toolbox. *NeuroImage*, *16*, 497.

Brett, L., & Willard, W. H. (2002). The origins of reciprocity and social exchange in friendships. *New Directions for Child and Adolescent Development, 95*, 27–40.

Brown J.W. & Braver T. S. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. *Science, 307*, 1118 –1121.

Bunge, S. A., Hazeltine, E., Scanlon, M. D., Rosen, A. C., & Gabrieli, J. D. (2002). Dissociable contributions of prefrontal and parietal cortices to response selection. *NeuroImage, 17*(3), 1562-1571.

Bunge, S. A., & Wright, S. B. (2007). Neurodevelopmental changes in working memory and cognitive control. *Current Opinion in Neurobiology, 17*, 243-250.

Burnett, S., & Blakemore, S. J. (2009). Functional connectivity during a social emotion task in adolescents and in adults. *European Journal of Neuroscience*, *29*, 1294–1301.

Buskens, V. (2003). Trust in triads: Effects of exit, control, and learning. *Games and Economic Behavior, 42*(2), 235–252.

C

Camerer, C. F. (2003). *Behavioral game theory*: Princeton University Press.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97*, 404-431.

Casey, B. J., Giedd, J. N., & Thomas, K. M. (2000). Structural and functional brain development and its relation to cognitive development. *Biological Psychology, 54*(1-3), 241-257.

Casey B.J., Jones R.M., & Hare T.A. (2008) The adolescent brain. *Annals of the New York Acadamy of Science*, *1124*, 111-126.

Casey, B. J., Davidson, M. C., Hara, Y., & Thomas, K. M. (2004). Early development of subcortical regions involved in non-cued attention switching. Developmental *Science, 7*(5), 534–542.

Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage, 12*(3), 314-325.

Choudhury, S., Blakemore, S. J., & Charman, T. (2006). Social cognitive development during adolescence. *Social Cognitive & Affective Neuroscience, 1*(3), 165-174.

Chang, L. J., B. B. Doll, M. van 't Wout, M. J. Frank & A. G. Sanfey Seeing is believing: Trustworthiness as a dynamic belief. *Cognitive Psychology, 61*(2), 87-105.

Chib, V. S., A. Rangel, S. Shimojo & J. P. O'Doherty (2009). Evidence for a Common Representation of Decision Values for Dissimilar Goods in Human Ventromedial Prefrontal Cortex. *Journal of Neuroscience,* 29(39), 12315-12320.

Cohen, M. X. (2008). Neurocomputational mechanisms of reinforcement-guided learning in humans: A review. *Cognitive, Affective, & Behavioral Neuroscience, 8*, 113-125.

Collins, W. A. (2003). More than myth: The developmental significance of romantic relationships during adolescence. *Journal of Research on Adolescence, 13,* 1–24.

Cools, R. (2008). Role of dopamine in the motivational and cognitive control of behavior. *The Neuroscientist, 14*(4), 381-395.

Corrado, G. & K. Doya (2007). "Understanding neural coding through the model-based analysis of decision making." *Journal of Neuroscience, 27*(31), 8178-8180.

Cosoco, C. A., Kollokian, V., Kwan, R. K. S., & Evans, A. C. (1997). Brainweb: Online interface of a 3-d mri simulated brain database. *NeuroImage, 5*(4), 425.

Crone, E. A., & van der Molen, M. W. (2004). Developmental changes in real life decision making: Performance on a gambling task previously shown to

depend on the ventromedial prefrontal cortex. *Developmental Neuropsychology, 25*(3), 251-279.

Crone E.A., Wendelken C., Donohue S., van Leijenhorst L., & Bunge S.A. (2006) Neurocognitive development of the ability to manipulate information in working memory. *Proceedings of the National Acadamy of Science USA*, *103*, 9315-9320.

Crone, E. A., Bullens, L., van der Plas, E. A. A., Kijkuit, E. J. & P. D. Zelazo (2008). Developmental changes and individual differences in risk and perspective taking in adolescence. *Development and Psychopathology, 20*(4), 1213-1229.

Crone, E. A., Zanolie, K., van Leijenhorst, L., Westenberg, P. M., & Rombouts, S. A. (2008b). Neural mechanisms supporting flexible performance adjustment during development. *Cognitive, Affective & Behavioral Neuroscience, 8*(2), 165-177.

Crone, E. A. (2009). Executive functions in adolescence: inferences from brain and behavior. *Developmental Science, 12*, 825-830.

D

Dale, A. M. (1999). Optimal experimental design for event-related fmri. *Human Brain Mapping, 8*(2), 109-114.

Davies P.L., Segalowitz S.J., & Gavin W.J. (2004) Development of response monitoring ERPs in 7- to 25-year-olds. *Developmental Neuropsychology, 25*, 355–376.

De Dreu, C. K. W. & van Lange, P. A. M. (1995). Impact of social value orientation on negotiator cognition and behavior. *Personality and Social Psychology Bulletin, 21*, 1177–1188.

De Vignemont, F., & Singer, T. (2006). The empathic brain: How, when and why? *Trends in Cognitive Sciences, 10*(10), 435-441.

Decety, J., Jackson, P. L., Sommerville, J. A., Chaminade, T., & Meltzoff, A. N. (2004). The neural bases of cooperation and competition: An fmri investigation. *NeuroImage, 23*(2), 744-751.

Decety, J., & Lamm, C. (2007). The role of the right parietal junction in social interaction: How low-level computational processes contribute to meta-cognition. *The Neuroscientist, 13*, 580-593.

Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience. 8*, 1611-1618.

Dosenbach, N. U., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences, 12*(3), 99-105.

Dumontheil, I., Apperly, I. A., & Blakemore, S.-J. (2009). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, 1-8.

E

Eisenberg, N., & Shell, R. (1986). The relation of prosocial moral judgment and behavior in children: The mediating role of cost. *Personality and Social Psychology Bulletin, 12*, 426–433.

Eisenberg, N., & Mussen, P. H. (1989). The Roots of Prosocial Behavior in Children. Cambridge: Cambridge University Press.

Eisenberg, N., Miller, P. A., Shell, R., McNalley, S., & Shea, C. (1991). Prosocial development in adolescence: A longitudinal study. *Developmental Psychology, 27*, 849–857.

Eisenberg, N., Carlo, G., Murphy, B., & Van Court, P. (1995). Prosocial development in late adolescence: A longitudinal study. *Child Development, 66*, 911–936.

Eisenberg, N., Guthrie I. K., et al. (2002). Prosocial development in early adulthood: a longitudinal study. *Journal of Personality and Social Psychology, 82*(6): 993.

Eisenberg N., Cumberland A., Guthrie I.K., & Murphy B.C. (2005). Age changes in prosocial responding and moral reasoning in adolescence and early adulthood. *Journal of Research on Adolescence*, *15*, 235-260.

Eisenberger, N.I., Gable, S.L., Lieberman, M.D. (2007). Functional magnetic resonance imaging responses relate to differences in real-world social experience. *Emotion, 7*,745-754.

Elkind D. (1985). Egocentrism redux. *Developmental Review*, *5*, 218-226.

Ernst, M., R. Romeo & S. Andersen (2008). Neurobiology of the development of motivated behaviors in adolescence window into a neural systems model, hormonal and molecular changes. *Pharmacology Biochemistry and Behavior, 93(3), 199-211.*

Estes, W. K. (1961). A descriptive approach to the dynamics of choice behavior. *Behavioral Science*, *6*, 177–184.

F

Fair, D. A., Cohen, A. L., Dosenbach, N. U., Church, J. A., Miezin, F. M., Barch, D. M., et al. (2008). The maturing architecture of the brain's default network. *Proceedings of the National Academy of Sciences USA, 105*(10), 4028-4032.

Fair, D. A., Cohen, A. L., Power, J. D., Dosenbach, N. U., Church, J. A., Miezin, F. M., et al. (2009). Functional brain networks develop from a

"local to distributed" organization. *PLoS Computationalt Biolology, 5*(5), e1000381.

Falk, A., Fehr, E., & Fischbacher, U. (2008). Testing theories of fairness—intentions matter. *Games and Economic Behavior*, *62, 1,* 287-303.

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior, 54*(2), 293-315.

Fehr, E., Bernhard, H., & Rockenbach, B. (2008) Egalitarianism in young children. *Nature*, 454(7208), 1079.

Fehr, E., & Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends in Cognitive Sciences, 11*, 419-427.

Fehr, E., & Gintis, H. (2007). Human motivation and social cooperation: Experimental and analytical foundations. *Annual Review of Sociology, 33*, 43-64.

Fehr, E., & Schmidt, K. M. (1999). A Theory Of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics*, *114*, 817-868.

Fletcher, P. C., Happé, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S. J., & Frith, C. D. (1995). Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. *Cognition, 57*(2), 109-128.

Forbes E.E., Dahl R.E. (2010) Pubertal development and behavior: Hormonal activation of social and motivational tendencies. *Brain and Cognition 72*, 66-72.

Frank, M. J., M. X. Cohen & A. G. Sanfey (2009). Multiple Systems in Decision Making: A Neurocomputational Perspective. *Current Directions in Psychological Science, 18*(2), 73-77

Frank, M. J., & Kong, L. (2008). Learning to avoid in older age. *Psychology and aging, 23*(2), 392-398.

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

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society London B Biological Sciences, 358*(1431), 459-473.

Frith, C. D., & Frith, U. (2007). Social Cognition in Humans. *Current Biology, 17*, 724-732.

Frith, C. D., & Frith, U. (2008). Implicit and Explicit Processes in Social Cognition. *Neuron, 60*, 503-510.

Frith, C. D. & T. Singer (2008). The role of social cognition in decision making. *Philosophical Transactions of the Royal Society B-Biological Sciences, 363*(1511), 3875-3886.

G

Gallagher, H. L., Jack, A. I., Roepstorff, A., & Frith, C. D. (2002). Imaging the intentional stance in a competitive game. *NeuroImage, 16*(3), 814-821.

Gallese, V. & Goldman, A. (1998) Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences, 12*, 493-501.

Galvan, A., Hare, T. A., Parra, C. E., Penn, J., Voss, H., Glover, G., et al. (2006). Earlier development of the accumbens relative to orbitofrontal cortex might underlie risk-taking behavior in adolescents. *Journal of Neuroscience, 26*(25), 6885-6892.

Galvan, A. (2010). Neural plasticity of development and learning. *Human Brain Mapping, 31*, 879-890.

Gardner, M. & Steinberg, L. (2005). Peer influence on risk taking, risk preference, and risky decision making in adolescence and adulthood: An experimental study. *Developmental Psychology*, *41*(4), 625-635.

Geier, C. F., Garver, K. E., & Luna, B. (2007). Circuitry underlying temporally extended spatial working memory. *NeuroImage, 35*(2), 904-915.

Giedd, J. N., J. Blumenthal, N. O. Jeffries, F. X. Castellanos, H. Liu, A. Zijdenbos, et al. (1999). Brain development during childhood and adolescence: a longitudinal MRI study. *Nature Neuroscience, 2*(10), 861-3.

Gogtay, N. & P. M. Thompson (2010). Mapping gray matter development: Implications for typical development and vulnerability to psychopathology. *Brain and Cognition, 72*(1), 6-15.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., et al. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences USA, 101*, 8174-8179.

Green, L., Fry, A.F., & Myerson, J. (1994) Discounting of delayed rewards: a life-span comparison. *Psychological Science 5*(1), 33–36.

Green L., Myerson J., & Ostaszewski P.(1999) Discounting of delayed rewards across the life span: age differences in individual discounting functions, *Behavioural Processes, 46*(1), 89-96.

Greenfield P.M., Keller H., Fuligni A., Maynard A. (2003). Cultural pathways through universal development. *Annual Review of Psychology 54*, 461-490.

Gummerum, M., Hanoch, Y., Keller, M. (2008). When child development meets economic game theory: an interdisciplinary approach to investigating social development. *Human Development, 51*(4),235-47

Gunther Moor, B., M. G. N. Bos, E. A. Crone & M. W. Van der Molen (2009). The heart-brake of social rejection: a heart rate analysis of developmetnal change in sensitivity to peer rejection *Psychophysiology, 46*, S125-S125.

Güroğlu, B., Haselager, G. J. T., van Lieshout, C. F. M., Takashima, A., Rijpkema, M., & Fernandez, G. (2008). Why are friends special? Implementing a social interaction simulation task to probe the neural correlates of friendship. *NeuroImage, 39*, 903-910.

Güroğlu, B., van den Bos, W., & Crone, E. A. (2009). Fairness considerations: Increasing understanding of intentionality in adolescence. *Journal of Experimental Child Psychology, 104*, 398-409.

Güroğlu, B., van den Bos, W., Rombouts, S. A. R. B., & Crone, E. A. (2010). Unfair? It depends: Neural correlates of fairness in social context. *Social Cognitive Affective Neuroscience, doi:10.1093/scan/nsq013*.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization, 3*, 367-388.

H

Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proceedings of the National Academy of Sciences of the United States of America, 105*, 6741-6746.

Hare, T. A., Camerer, C. F., Knoepfle, D. T., O'Doherty, J. P., & Rangel, A. (2010). Value Computations in Ventral Medial Prefrontal Cortex during Charitable Decision Making Incorporate Input from Regions Involved in Social Cognition. *Journal of Neuroscience, 30*, 583-590

Harris, L. T., McClure, S., van den Bos, W., Cohen, J. D., Fiske, S. T. (2007). Regions of MPFC differentially tuned to social and nonsocial affective stimuli. *Cognitive, Affective, & Behavioral Neuroscience*, *7*, 309-316.

Hartup, W. W. & N. Stevens (1997). Friendships and adaptation in the life course. *Psychological Bulletin, 121*(3), 355-370.

Hester, R., J. Madeley, K. Murphy & J. B. Mattingley (2009). Learning from errors: error-related neural activity predicts improvements in future inhibitory control performance. *Journal of Neuroscience, 29*(22), 7158-65.

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: reinforcement learning, dopamine and the error-related negativity. *Psychological Review, 109*, 679-709.

Honey C.J., Kitter R., Breakspear M., Sporns O. (2007). Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proceedings of the National Academy of Sciences USA, 104*, 10240-10245.

Honey C.J., Sporns O., Cammoun L., Gigandet X., Thiran J.P., Meuli R., et al. (2009). Predicting human resting-state functional connectivity from

structural connectivity. *Proceedings of the National Academy of Sciences USA, 106,* 2035-2040.

Hooper, C. J., Luciana, M., Conklin, H. M., & Yarger, R. S. (2004). Adolescents' performance on the iowa gambling task: Implications for the development of decision making and ventromedial prefrontal cortex. *Developmental Psychology, 40*(6), 1148-1158.

Huizinga, M., C. V. Dolan & M. W. van der Molen (2006). Age related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia, 44*(11), 2017-2036.

Huttenlocher, P. R. (1979). Synaptic density in human frontal cortex-developmental changesn and effects of aging. *Brain Research, 163*(2), 195-205.

I

Izuma K., Saito D. N., & Sadato N. (2008) Processing of Social and Monetrary Rewards in the Human Striatum. *Neuron, 58,* 284-294.

J

Johnson, M.H. (2001). Functional brain development in humans. *Nature Reviews of Neuroscience*, *2*, 475-483.

Johnson, M.H. (2005). *Developmental cognitive neuroscience(2nd edition).* Oxford, Blackwell.

Johnson, M.H. (2011). Interactive Specialization: A domain-general framework for human functional brain development? *Developmental Cognitive Neuroscience, 1*, 7-21.

Johnson, M.H., Grossmann, T., Kadosh, K. (2009). Mapping functional brain development: Building a social brain through interactive specialization. *Developmental Psychology, 45,* 151-159.

K

Kelly, A. M., Di Martino, A., Uddin, L. Q., & Shehzad, Z. (2008). Development of anterior cingulate functional connectivity from late childhood to early adulthood. *Cerebral Cortex, 19* (3), 640-657.

Kerns, J. G. (2006). Anterior cingulate and prefrontal cortex activity in an fmri study of trial-to-trial adjustments on the simon task. *NeuroImage, 33*(1), 399-405.

Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science, 303*(5660), 1023-1026.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P.R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science, 308*(5718), 78-83.

King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., & Montague, P. R. (2008). The rupture and repair of cooperation in borderline personality disorder. *Science, 321*(5890), 806-810.

Kishida K.T., King-Casas B., Montague P.R. (2010) Neuroeconomic Approaches to Mental Disorders. *Neuron, 67*, 543-554.

Klein, T. A., Neumann, J., Reuter, M., Hennig, J., von Cramon, D. Y., & Ullsperger, M. (2007). Genetically determined differences in learning from errors. *Science, 31*8(5856), 1642-1645.

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Increased brain activity in frontal and parietal cortex underlies the development of visuospatial working memory capacity during childhood. *Journal of Cognitive Neuroscience, 14*(1), 1-10.

Knoch, D., Gianotti, L. R. R., Pascual-Leone, A., Treyer, V., Regard, M., Hohmann, M., et al. (2006). Disruption of Right Prefrontal Cortex by Low-Frequency Repetitive Transcranial Magnetic Stimulation Induces Risk-Taking Behavior. *Journal of Neuroscience, 26*, 6469-6472.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science, 314*(5800), 829-832.

Knoch, D., M. A. Nitsche, U. Fischbacher, C. Eisenegger, A. Pascual-Leone & E. Fehr (2008). Studying the neurobiology of social interaction with transcranial direct current stimulation--the example of punishing unfairness. *Cereb Cortex* 18(9), 1987-90.

Kohlberg, L. (1981). *The meaning and measurement of moral development*. Worcester, MA: Clark University Press.

Kopp, C. B. (1982). Antecedents of self-regulation: A developmental perspective. *Developmental Psychology, 18*, 199–214.

Kramer, R. M., McClintock, C. G., & Messick, D. M. (1986). Social values and cooperative response to a simulated resource conservation crisis. *Journal of Personality, 54*(3), 576-582.

Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., Heinecke, A., & Grafman, J. (2007). Neural correlates of trust. *Proceedings of the National Acadamy Science USA, 104*(50), 20084-20089.

Krueger F, Grafman J, McCabe K. (2008). Neural correlates of economic game playing. *Philosophical Transactions of the Royal Society London B Biological Sciences, 363*, 3859-3874.

L

Ladouceur, C. D., Dahl, R. E., & Carter, C. S. (2004). ERP correlates of action monitoring in adolescence. In R. E. Dahl & L. P. Spear (Eds.), Adolescent brain development: Vulnerabilities and opportunities (Annals of the New York Academy of Sciences, Vol. 1021, pp. 329-336). New York: New York Academy of Sciences.

Lahno, B. (1995). Trust, reputation, and exit in exchange relationships. *Journal of Conflict Resolution, 39*(3), 495-510.

Lamm, C., Batson, C. D., & Decety, J. (2007). The neural substrate of human empathy: Effects of perspective-taking and cognitive appraisal, *Journal of Cognitive Neuroscience, 19*, 42-58.

Lamm, C., & Singer, T. (2010). The role of anterior insular cortex in social emotions. *Brain Structure and Function, 214*, 579-591.

Luna, B., & Sweeney, J. A. (2001). Studies of brain and cognitive maturation through childhood and adolescence: A strategy for testing neurodevelopmental hypotheses. *Schizophrenia Bulletin, 27*(3), 443-455.

Lieberman M.D. (2007). Social cognitive neuroscience, A review of core processes. *Annual Review of Psychology*, *58*, 259-289.

M

Malhotra, D. (2004). Trust and reciprocity decisions: The differing perspectives of trustors and trusted parties. *Organizational Behavior and Human Decision Processes, 94*(2), 61-73.

Mareschal et al., (2007) D. Mareschal, M.H. Johnson, S. Sirois, M. Spratling, M. Thomas and G. Westermann, Neuroconstructivism, Vol. I: How the brain Constructs Cognition, Oxford University Press, Oxford.

Martin, J., Sokol, B. W., & Elfers, T. (2008). Taking and coordinating perspectives: From prereflective interactivity, through reflective intersubjectivity, to metareflective sociality. *Human Development, 51*(5-6), 294.

McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedingsof the National Academy of Sciences USA, 98*(20), 11832-11835.

McClintock, C. G., & Allison, S. T. (1989). Social value orientation and helping behavior. *Journal of Applied Social Psychology, 19*(4), 353-362.

McClure-Tone, Fromm, S., Blair, R. J., Pine, D. S., & Ernst, M. (2008). Amygdala and nucleus accumbens activation to emotional facial

expressions in children and adolescents at risk for major depression. *American Journal of Psychiatry, 165*(2), 266.

Miller E.K., & Cohen J.D. (2001). An integerative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167-202.

Mitchell, J. P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cerebral Cortex, 18*(2), 262-271.

Montague, P.R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science, 308*(5718), 78-83.

Montague, P. R., & Lohrenz, T. (2007). To detect and correct: Norm violations and their enforcement. *Neuron, 56*(1), 14.

Munakata Y., McClelland J.L. 2003. Connectionist models of development. *Developmental Science, 6*, 413-429.

N

Nelson, E. E., E. Leibenluft, E. B. McClure & D. S. Pine (2005). The social re-orientation of adolescence: a neuroscience perspective on the process and its relation to psychopathology. *Psychological Medicine, 35*(2), 163-174.

Northoff, G., Heinzel, A., de Greck, M., & Bermpohl, F. (2006). Self-referential processing in our brain—a meta-analysis of imaging studies on the self. *NeuroImage, 31*(1), 440-457.

O

O'Brien, S. F. & Bierman, K. L. (1988) Conceptions and perceived influence of peer groups: Interviews with preadolescents and adolescents. *Child Development*, 59(5), 1360-1365

Ochsner, K. N. (2008). The social-emotional processing stream: Five core constructs and their translational potential for schizophrenia and beyond. *Biological Psychiatry, 64*(1), 48-61.

Olson, L. A., T. A. Oberndorfer, T. T. Yang & G. K. Frank (2008). Reactive aggressive youth brain activation is increased in response to emotional faces, but reduced during an impulsivity task. *Biological Psychiatry, 63*(7), 47.

P

Pagnoni, G., C. F. Zink, P. R. Montague & G. S. Berns (2002). Activity in human ventral striatum locked to errors of reward prediction. *Nature Neuroscience, 5*(2), 97-98.

Paus, T. (2010). Growth of white matter in the adolescent brain: Myelin or axon? *Brain and Cognition, 72*(1), 26-35.

Paus T. (2010) Population neuroscience: why and how. *Human Brain Mapping, 31,* 891-903.

Paus, T., D. L. Collins, A. C. Evans, G. Leonard & B. Pike (2001). Maturation of white matter in the human brain: a review of magnetic resonance studies. *Brain Research Bulletin, 54*(3), 255-266.

Paus, T., R. Toro, G. Leonard, J. V. Lerner, R. M. Lerner, M. Perron, et al. (2008). Morphological properties of the action-observation cortical network in adolescents with low and high resistance to peer influence. *Social Neuroscience, 3*(3-4), 303-316.

Pfeifer, J. H., M. D. Lieberman & M. Dapretto (2007). "I Know You Are But What Am I?!": Neural Bases of Self-and Social Knowledge Retrieval in Children and Adults. *Journal of cognitive neuroscience* 19(8), 1323-37.Paus T., Keshavan M., & Giedd J.N. (2008). Why do many psychiatric disorders emerge during adolescence? *Nature Reviews of Neuroscienc*e, *9*, 947-957.

Pfeifer J.H., Lieberman M.D., & Dapretto M. (2007). "I know you are but what am I?!", neural bases of self- and social knowledge retrieval in children and adults. *Journal of Cognitive Neuroscience*, *19*, 1323-1337.

Phan, K. L., Sripada, C. S., Angstadt, M., McCabe, K. (2010). Reputation for reciprocity engages the brain reward center. *Proceedings of the National Academy of Sciences USA, 107*, 13099-13104.

Piaget, J. (1956). *The origins of intelligence in the child*. New York: International Universities Press.

Pillutla, M., Malhotra, D., & Murnighan, K. J. (2003). Attributions of trust and the calculus of reciprocity. *Journal of Experimental Social Psychology. 39*(5), 448-455.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences, 10*(2), 59.

Poldrack R.A. (2010) Interpreting developmental changes in neuroimaging signals. *Human brain mapping, 31,* 872-878.

Preuschoff, K., Quartz, S. R., & Bossaerts, P. (2008). Human insula activation reflects risk prediction errors as well as risk. *Journal of Neuroscience*, *28*(11), 2745.

Q

Qu L., & Zelazo P.D. (2007) The facilitative effect of positive stimuli on 3-year-olds' flexible rule use. *Cognitive Development, 22*, 456-473.

R

Rangel, A., C. Camerer & P. Montague (2008). A framework for studying the neurobiology of value-based decision making. *Nature Review of Neuroscience, 9*(7), 545-556.

Raven J.C. (1941). Standardisation of progressive matrices. *British Journal of Medical Psychiatry, 19*, 137-150.

Raven, J., Raven, J. C., & Court, J. H. (1998, updated 2003). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. Section 1: General Overview. San Antonio, TX: Harcourt Assessment

Rescorla, R. A. & A. R. Wagner (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Classical Conditioning II. A. H. Black and W. F. Prokasy, Appleton-Century-Crofts**:** 64-99.

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science, 306*(5695), 443-447.

Ridderinkhof, K. R., & van den Wildenberg, W. P. (2005). Neuroscience. Adaptive coding. *Science, 307*(5712), 1059-1060.

Rilling, J., B. King-Casas & A. G. Sanfey (2008). The neurobiology of social decision-making. *Current Opinion in Neurobioly* 18 (2), 159-65.

Rilling, J. K., A. G. Sanfey, J. A. Aronson, L. E. Nystrom & J. D. Cohen (2004). The neural correlates of theory of mind within interpersonal interactions. *NeuroImage* 22(4), 1694.

Rilling, J., Glenn, A., Jairam, M., Pagnoni, G., Goldsmith, D., Elfenbein, H., & Lilienfeld, S. (2007). Neural correlates of social cooperation and non-cooperation as a function of psychopathy. *Biological Psychiatry*, *61*(11), 1260-1271.

Rilling, J.K., Goldsmith, D., Glenn, A., & Jairam, M. (2008). The neural correlates of the affective response to unreciprocated cooperation. *Neuropsychologia, 46*(5), 1256-66.

Rilling J.K., Sanfey A.G. (2010). The Neuroscience of Social Decision-Making. *Annual Review of Psychology*. (e-pub ahead of print)

Rotenberg, K.J. (1995). The socialization of trust: Parents' and children's interpersonal trust. *International Journal of Behavioral Development, 18,* 713-726.

Rotenberg, K.J., Fox, C., Green, S., Ruderman, L., Slater, K., Stevens, K., Carlo, S. (2005). Construction and validation of a children's interpersonal trust belief scale. *British Journal of Developmental Psychology, 23,* 271-292.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23*(3), 393-404.

Rushworth, M. F., & Behrens, T. E. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. Nature Neuroscience, 11(4), 389-397.Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., & Kilts, C. (2002). A neural basis for social cooperation. *Neuron, 35*(2), 395-405.

Rushworth, M. F. S. & T. E. J. Behrens (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nature Neuroscience, 11*(4), 389-397.

S

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science, 300*(5626), 1755-1758.

Sanfey, A. G. (2007). Social Decision-Making: Insights from Game Theory and Neuroscience. *Science,* 318(5850), 598-602.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *NeuroImage, 19*(4), 1835-1842.

Saxe, R., S. Carey & N. Kanwisher (2004). Understanding Other Minds: Linking Developmental Psychology and Functional Neuroimaging. *Annual Review of Psychology,* 55, 87-124.

Sebastian, C., Burnett, S., & Blakemore, S.-J. (2008). Development of the self-concept during adolescence. *Trends in Cognitive Sciences, 12*, 441-446.

Selman, R.L. (1980). *The growth of interpersonal understanding*. New York: Academic Press.

Schaffer, H. R. (1996). *Social Development*. UK: Blackwell Publishing.

Schmithorst, V. J. & W. H. Yuan (2010). White matter development during adolescence as shown by diffusion MRI. *Brain and Cognition, 72*(1), 16-25.

Schultz, W. (2007). Behavioral dopamine signals. *Trends in Neurosciences, 30*(5), 204-210.

Sebastian, C., S. Burnett & S. J. Blakemore (2008). Development of the self-concept during adolescence. *Trends in Cognitive Sciences, 12*(11), 441-446.

Shanks, D. R., Tunney, R. J., & McCarthy, J. D. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making, 15*(3), 233-250.

Shaw P., Kabani N.J., Lerch J.P., Eckstrand K., Lenroot R., Gogtay N., Greenstein D., Clasen L., Evans A., Rapoport J.L., Giedd J.N., & Wise S.P. (2008). Neurodevelopmental trajectories of the human cerebral cortex. *Journal of Neuroscience*, *28*, 3586-3594.

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature, 439*(7075), 466-469.

Singer, T., Critchley, H. D., & Preuschoff, K. (2009). A common role of insula in feelings, empathy and uncertainty. *Trends in Cognitive Sciences, 13*, 334-340.

Singer, T., & Steinbeis, N. (2009). Differential Roles of Fairness- and Compassion-Based Motivations for Cooperation, Defection, and Punishment. *Annals of the New York Academy of Sciences, 1167*, 41-50.

Somerville, L. H. & B. J. Casey (2010). Developmental neurobiology of cognitive control and motivational systems. *Current Opinion in Neurobiology* 20(2), 236-241.

Somsen, R. J. (2007). The development of attention regulation in the wisconsin card sorting task. *Developmental Science, 10*(5), 664-680.

Spear L.P. (2000). The adolescent brain and age-related behavioral manifestations. *Neuroscience and Biobehavioral Review, 24*, 417-463.

Spitzer, M., Fischbacher, U., Herrnberger, B., Grön, G., & Fehr, E. (2007). The neural signature of social norm compliance. *Neuron, 56*(1), 185-196.

Sripada CS, Angstadt M, Banks S, Nathan PJ, Liberzon I, Phan KL. (2009). Functional neuroimaging of mentalizing during trust in social anxiety disorder. *Neuroreport, 20*, 984-949.

Steinberg, L. (2005). Cognitive and affective development in adolescence. *Trends In Cognitive Science, 9*(2), 69-74.

Steinberg, L. & K. C. Monahan (2007). Age differences in resistance to peer influence. *Developmental Psychology, 43*(6), 1531-1543.

Steinberg L. (2009). Adolescent development and juvenile justice. *Annual Review of Clinical Psychology*, *5*, 459-485.

Steinberg L., Graham S., O'Brien L., Woolard J., Cauffman E., Banich M. (2009). Age differences in future orientation and delay discounting. *Child Development, 80,* 28-44.

Sterzer, P., C. Stadler, F. Poustka & A. Kleinschmidt (2007). A structural neural deficit in adolescents with conduct disorder and its association with lack of empathy. *NeuroImage, 37*(1), 335-342.

Straub, P. G. & J. K. Murnighan (1995). An experimental investigation of ultimatum games - Information, fairness, expectations and lowest acceptable offers. *Journal of Economic Behavior and Organization, 27*(3), 345-364.

Supekar, K., M. Musen & V. Menon (2009). Development of Large-Scale Functional Brain Networks in Children. *Plos Biology,* 7(7).

Sutter, M., & Kocher, M. G. (2007). Trust and trustworthiness across different age groups. *Games and Economic Behavior, 59*(2), 364-382.

Sutter, M. (2007). Outcomes versus intentions: On the nature of fair behavior and its development with age. *Journal of Economic Psychology* 28(1), 69-78.

Sutton, R. S. & A. G. Barto (1999). Reinforcement learning. *J Cog Neurosci*.

T

Tabibnia, G., Satpute, A. B., & Lieberman, M. D. (2008). The sunny side of fairness: Preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychological Science, 19*(4), 339-347.

Takahashi, H., Yahata, N., Koeda, M., Matsuda, T., Asai, K., & Okubo, Y. (2004). Brain activation associated with evaluative processes of guilt and embarrassment: an fMRI study. *NeuroImage, 23*, 967-974.

Talairach, J., & Tourneaux, P. (1988). Co-planar stereotaxic atlas of the human brain. Stuttgart: Thieme Verlag.

Tankersley, D., Stowe, C. J., & Huettel, S. A. (2007). Altruism is associated with an increased neural response to agency. *Nature Neuroscience, 10*, 150-151.

Tanner J.M. (1975). The measurement of maturity. *Transactions European Orthodontic Society*, *45*, 42-56.

Taylor, S. F., Stern, E. R., & Gehring, W. J. (2007). Neural systems for error monitoring: Recent findings and theoretical perspectives. *The Neuroscientist, 13*(2), 160-172.

Toga, A. W., P. M. Thompson & E. R. Sowell (2006). Mapping brain maturation. *Trends in Neurosciences, 29*(3), 148-159.

Toni, I., Rowe, J., Stephan, K. E., & Passingham, R. E. (2002). Changes of cortico-striatal effective connectivity during visuomotor learning. *Cerebral Cortex,. 12(10), 1040-1047.*

U

Underwood, B., & Moore, B. (1982). Perspective-taking and altruism. *Psychological Bulletin, 91*, 143–173.

V

van den Bos, W., McClure, S. M., Harris, L. T., Fiske, S. T., & Cohen, J. D. (2007). Dissociating affective evaluation and social cognitive processes in the ventral medial prefrontal cortex. *Cognitive, Affective, & Behavioral Neuroscience, 7*(4), 337-346.

van den Bos W., Westenberg M., Van Dijk E., Crone E.A. (2010) Development of trust and reciprocity in adolescence. *Cognitive Development, 25*, 90-102.

van den Bos W., van Dijk E., Westenberg M., Rombouts S.A., & Crone E.A. (2009). What motivates repayment? Neural correlates of reciprocity in the Trust Game. *Social Cognitive & Affective Neuroscience, 4*, 294-304.

van den Bos, W., B. Güroğlu, B. G. van den Bulk, S. Rombouts & E. A. Crone (2009). Better than expected or as bad as you thought? The neurocognitive development of probabilistic feedback processing. *Frontiers in human neuroscience* 3.

van den Bos, W., E. Van Dijk, P. M. Westenberg, S. A. R. B. Rombouts, E. A. Crone & (2010). Changing Brains, Changing Perspectives: The Neurocognitive Development of Reciprocity. *Psychological Science*.

van Duijvenvoorde, A. C., Zanolie, K., Rombouts, S. A., Raijmakers, M. E., & Crone, E. A. (2008). Evaluating the negative or valuing the positive? Neural mechanisms supporting feedback-based learning across development. *Journal of Neuroscience, 28*(38), 9495-9503.

van Lange, P. A. M., Otten, W., de Bruin, E. M. N., & Joireman, J. A. (1997). Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. *Journal of Personality and Social Psychology, 73*(4), 733-746.

van Lange, P. A. M. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology, 77*(2), 337-349.

van Leijenhorst, L., Zanolie, K., Van Meel, C. S., Westenberg, P. M., Rombouts, S. A., & Crone, E. A. (2009). What motivates the adolescent? Brain regions mediating reward sensitivity across adolescence. *Cerebral Cortex 20(1), 61-69.*

van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping, 30*, 829-858.

van 't Wout, M., Kahn, R. S., Sanfey, A. G., & Aleman, A. (2005). Repetitive transcranial magnetic stimulation over the right dorsolateral prefrontal cortex affects strategic decision-making. *NeuroReport, 16*, 1849-1852.

Velanova, K., Wheeler, M. E., & Luna, B. (2008). Maturational changes in anterior cingulate and frontoparietal recruitment support the development

of error processing and inhibitory control. *Cerebral Cortex, 18*(11), 2505-2522.

W

Wang A.T., Lee S.S., Sigman M., & Dapretto M. (2006). Developmental changes in the neural basis of interpreting communicative intent. *Social Cognitive & Affective Neuroscience, 1*, 107-121.

Ward B.D. (2000). Simultaneous inference for fMRI data. Available at: http://afni.nimh.nih.gov/afni/docpdf/AlphaSim.pdf. Accessed January 10, 2009

Wechsler, D. (1991). Wechsler intelligence scale for children-third edition. Manual. San Antonio: The Psychological Corporation.

Wechsler, D. (1997). Wechsler adult intelligence scale—third edition. Administration and scoring manual. San Antonio: The Psychological Corporation.

X

-

Y

Yamagishi, T., Cook, K. S., & Watabe, M. (1998). Uncertainty, trust, and commitment formation in the United States and Japan. *American Journal of Sociology*, *104,*165–194.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review, 111*(4), 931-959.

Youniss, J. 1980. *Parents and Peers in Social Development: A Sullivan-Piaget Perspective*. Chicago: University of Chicago Press.

Z

Zanolie, K., Van Leijenhorst, L., Rombouts, S. A., & Crone, E. A. (2008). Separable neural mechanisms contribute to feedback processing in a rule-learning task. *Neuropsychologia, 46*(1), 117-126.

Zimmerman, B. J. (2000). Attaining Self-Regulation: A Social Cognitive Perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner, Handbook of Self-Regulation. San Diego: Academic Press.

CV

Wouter van den Bos was born on May 23rd 1980 in Amsterdam, the Netherlands. He graduated from the Werkplaats Kindergemeenschap at Bilthoven in 1998. Subsequently, he has received his M.A. in Philosophy (2005) and M.Sc. in Cognitive Science (2006), both at the University of Amsterdam. For the Cognitive Science Research Master he spent 8 months as a visiting researcher at Princeton University, in the laboratory of prof. dr. Jonathan Cohen. In 2006 he started his PhD program at Leiden University, with prof. dr. Eveline Crone, prof. dr. Eric van Dijk and prof. dr. Michiel Westernberg as advisors. In February Wouter joined Stanford University's department of psychology to work as a postdoctoral researcher in Dr. Samual McClure's Decision Neuroscience Lab.