BAYESIAN HIERARCHICAL PREDICTIVE CODING OF HUMAN SOCIAL BEHAVIOUR

HAUKE FRERK HILLEBRANDT

DISSERTATION SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

OF THE

UNIVERSITY COLLEGE LONDON

INSTITUTE OF COGNITIVE NEUROSCIENCE

UNIVERSITY COLLEGE LONDON

Declaration

I, Hauke Frerk Hillebrandt, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Hillebrandt

Acknowledgements

According to Matt Ridley, nobody on the planet actually really knows how to make something as simple as a computer mouse. Some people might know how all the parts in a computer mouse work and can put them together, but no one person knows how to build the oil rig to extract the oil needed for the plastic for the mouse, make the soap that keeps the people on the oil rig clean, etc. The point being: modern technology is made possible by division of labour and specialization. Similarly, science does not only stand on the shoulders of giants (i.e. other scientists), but also on the labour of billions of people who make modern research possible – including my PhD research. In this sense, I would like thank everyone and acknowledge everyone's contribution to this research and hope those people I forget to thank here specifically will forgive me :-)

First, I would like to thank Sarah-Jayne Blakemore for being an outstanding enabler, role model and mentor, who went beyond the call of duty in her role as an advisor. Without her encouragement, advice and faith in me, this thesis would not have been possible. I would also like to thank: Jonathan Roiser, my second supervisor, who was always very generous with his time and helpful when teaching me neuroimaging analysis, Iroise Dumontheil, for being a great collaborator and unofficial supervisor who taught me a lot about research, Joshua Greene, my supervisor during the final years of my PhD at Harvard, who welcomed me in his lab and has been very supportive of my research, Uta Frith, for mentorship and helpful discussions, Karl Friston, for mentorship and valuable discussion, Kathrin Cohen Kadosh, for teaching me about SPM and fMRI, the people of the Developmental lab at the Institute of Cognitive Neuroscience, the Moral Cognition lab at Harvard, and my PhD cohorts at UCL and Harvard, especially, P., Elisa, Kate, Nahid, Tobias, Slava, Micah, David, Joe, Alek, Steven, Elinor, and Sara. I would also like to thank my early mentor Arvid Kappas introducing me to social neuroscience.

I am also very grateful to for valuable discussion and comments on drafts of manuscripts that the various chapters of this thesis are based on: Pradheep Shanmugalingam for comments on chapter 1, Nick Wright for early discussion of the experimental design of the experiment described in Chapter 2, Karl Friston for discussion and comments on the manuscript that chapter 4 is based on, and Uta Frith, Chris Frith, James Kilner and Micah Allen comments on chapter 5.

I would also like to thank the admistrative and technical staff at UCL and Harvard, especially to Lambert Dean, for being an amazing supervisor for my IT demonstratorship, and the people of ICN Tech team, for being great colleagues, the BUCNI neuroimaging team, in particular Oliver Josephs, for helpful discussions, and Monika Glinska, for being very helpful with paperwork.

I am also grateful for funding from University College London in form of a scholarship, and the European Commission's INCORE grants as well as the Charlotte and Yule Bogue fellowship that made my visit to Harvard possible.

Finally, I would like to thank my family and friends, especially Karin, Hermann, Ines and Lauren for their support.

Thesis Abstract

'Bayesian hierarchical predictive coding of human social behaviour.'

Biological agents are the most complex systems humans encounter in their natural environment and it is critical to model other's mental states correctly to predict their behaviour. To do this one has to generate a mental representation based on an internal neural model of the other agent (Chapter 1). Here we show, in a series experiments, that people use and update their Bayesian priors in social situations and explain how they create mental representations of others to guide action selection. We investigate the neural mechanisms and the brain connectivity that underlie these social processes and how they develop with age.

In chapter 2, we show how experimentally induced prior experience with other people (here social inclusion or exclusion) influences the level of trust towards those people.

In chapter 3, we describe an fMRI study using a social perspective-taking task that examines the developmental differences between adolescents and adults in the control of action selection by social information. Using the same task, in chapter 4, we investigate the effective connectivity between the activated regions with Dynamic causal modelling.

In Chapter 5, we explore effective connectivity of fMRI data from the Human connectome project (Van Essen et al., 2012). During the task participants viewed animations of triangles moving either randomly or so that they evoke mental state attribution (Castelli et al., 2000).

Chapter 6 concludes with a summary of the experiments and integrate them into existing research, as well as provide a critical synthesis of the findings in order to suggest future research directions. We interpret our findings in a hierarchical predictive coding framework, where agents

try to create a neural model of the external world to minimize prediction errors, Bayesian surprise and free energy.

Contributions

The work presented in this PhD thesis would not have been possible without the help of my collaborators:

Chapter 2:

Hillebrandt, H., Sebastian, C. & Blakemore, S. (2011). Experimentally induced social inclusion influences behavior on trust games. *Cognitive Neuroscience*, 2(1), 27-33. doi:10.1080/17588928.2010.515020

My contribution to this study: I conceived the specifics of the study design and programmed the experiment. I recruited and tested all the participants. I analysed all the data. I wrote the first draft of the paper and wrote the final draft of the paper collaboratively with the other authors. I implemented all the reviewers comments.

Chapter 3:

Hillebrandt, H.*, Dumontheil, I.*, Apperly, I. A., & Blakemore, S.-J. (2012). Developmental Differences in the Control of Action Selection by Social Information. *Journal of Cognitive Neuroscience*, 24(10), 2080–2095. doi:10.1162/jocn_a_00268 * indicates shared first authorship

My contribution to this study: The adult participants were recruited and scanned by Iroise Dumontheil. I recruited all the adolescent participants. I scanned the adolescent participants together with Iroise Dumontheil. Both Iroise Dumontheil and I analysed the neuroimaging data independently. I interpreted the data and wrote parts of the first draft of the paper with Iroise

Dumontheil. I wrote the final version of the paper collaboratively with the other authors. I was involved in implementing reviewer's comments.

Chapter 4:

Hillebrandt, H., Dumontheil I., Blakemore, S, Roiser, J.P. (2013). Dynamic Causal Modelling of effective connectivity during perspective taking in a communicative task *Neuroimage*. doi.org/10.1016/j.neuroimage.2013.02.072

My contribution to this study: The adult participants were recruited and scanned by Iroise Dumontheil. I analysed and interpreted all the data as described in the chapter. I wrote the first draft of the paper and wrote the final draft of the paper collaboratively with the other authors. I implemented the reviewer's comments.

Chapter 5:

Hillebrandt, H., Friston, K.J., Blakemore, S-J. (in press). Effective connectivity during animacy perception - dynamic causal modelling of Human Connectome Project data. *Scientific Reports*

My contribution to this study: the Human Connectome Project Consortium recruited and scanned all the participants. I analysed and interpreted all the data. I wrote the first draft of the paper and wrote the final draft of the paper collaboratively with the other authors. I implemented the reviewer's comments.

I would like to thank all my collaborators and will use the 1st person plural throughout the experimental chapters of the thesis to acknowledge their contributions.

Other work during the PhD

During my PhD, I worked on other issues not mentioned in this thesis, which have resulted in the following publication:

Bazargani N., Hillebrandt, H., Christoff, K, Dumontheil, I. (2013). Developmental changes in effective connectivity associated with relational reasoning. Human Brain Mapping

Uncited Sources that this partly PhD is based on:

Bazargani N., Hillebrandt, H., Christoff, K, Dumontheil, I. (2013). Developmental changes in effective connectivity associated with relational reasoning. Human Brain Mapping

Hillebrandt. H. *Unpublished Coursework* 2009, 2010

Hillebrandt, H. *Pubpeer comment* 2013

Contents

1. Introduction

1.1. Definitions and theoretical considerations about social neuroscience

Humans are a profoundly social species and much of human behaviour relates to our unique sociality (Frith & Frith, 2001). Social cognition has been defined as the act of acquiring knowledge about the mental states of others, like their beliefs, desires, and intentions as well as the meaning of what other people communicate (Przyrembel, Smallwood, & Singer, 2012). Others have argued that the term "social cognition" can be used to describe all socially relevant processes “including action intention understanding, affective resonance and empathy, face recognition, social memory, and many others” (Przyrembel, Smallwood, & Singer, 2012). “Theory of mind” (ToM) or “mentalizing” refers to the ability to make inferences about mental states such as beliefs, desires and intentions (Frith & Frith, 2007). Usually, mentalizing specifically refers to cognitive perspective taking processes and the underlying ToM brain network (Przyrembel et al., 2012). The ability to infer mental states enables humans to understand and predict other people’s behaviour. There has been extensive investigation of the development of ToM in children (Carpenter, Nagell, & Tomasello, 1998). The understanding of mental states develops in a step-wise-fashion during the first 4 or 5 years of life (Frith and Frith, 2003).

Social neuroscience is a field that has developed a lot since the 1990s, in part because of innovations in neuroimaging (Brothers, 1992, Frith and Frith, 2010). It is now possible to image the living human brain of a participant taking part in a pyschology experiment, and to do so with good temporal and/or spatial resolutions using functional Magnetic Resonance Imaging (fMRI), Positron Emission Tomography (PET), magnetoencephalography (MEG), electroencephalography (EEG), or near infrared spectroscopy (NIRS). Social neuroscience aims to investigate the neural instantiations of basic and advanced social perception, processing, and cognition, either directly through brain imaging or indirectly by inferring mechanisms of social processes through behaviour.

In this thesis, I use fMRI to study social brain function, and so I will now briefly describe the method and its main assumptions and limitations.

Functional magnetic resonance imaging (fMRI) is a technique to measure properties of blood flowing through a living animal (in this case, in the brain). fMRI is often used to measure the blood oxygenation level dependent (BOLD) signal in the human brain. The BOLD signal measures changes in the levels of oxygenated and deoxygenated haemoglobin, which have different magnetic properties (Ogawa, Lee, Kay, & Tank, 1990). In task-based cognitive neuroscience experiments, participants are often exposed to different experimental stimuli, which trigger different sensations, perceptions, cognitions or mental states, while measuring the BOLD signal. When subtracting the brain images acquired during the exposure to experimental stimuli from those acquired during a control task, a map of brain can be obtained that shows local increases in cerebral blood flow (CBF) (Harris, Reynell, & Attwell, 2011). Increases in CBF within a brain region are associated with neural activity (Harris et al., 2011) and so BOLD signal is commonly interpreted as proxy for neural activity. In other words, the difference between the brain maps can be used to localize those areas of a brain that are relatively more active during the task of interest relative to the control condition by means of tracking blood flow to those areas.

The assumption that the BOLD signal is reflecting neural activity in the sense of action potentials in that particular area is controversial and which part of neural activity exactly the BOLD signal reflects is part of an ongoing discussion (Harris et al., 2011). Moreover, because fMRI measures blood flow and changes in the blood flow are in the order of seconds unlike neuronal action potentials, which are faster, it is generally agreed that fMRI has poor temporal resolution on single experimental trials. However, with fMRI the whole brain can be scanned and the spatial resolution is getting increasingly better (Logothetis, 2008).

Finally, the BOLD signal cannot easily differentiate between function-specific processing and neuromodulation and might confuse excitation and inhibition (Logothetis, 2008).

1.1.1. Defining Social Neuroscience

First, what research constitutes social neuroscience must be defined as this will be important to contextualize the experimental chapters in this thesis. The founding editors of the journal 'Social Neuroscience,' Jean Decety and Julian Paul Keenan, have proposed a working definition of social neuroscience, which they say should be treated as an inclusive, rather than an exclusive, guide. They write:

> Social neuroscience may be broadly defined as the exploration of the neurological underpinnings of the processes traditionally examined by, but not limited to, social psychology. This broad description provides a starting point from which we may examine the neuroscience of social behavior and cognition. (Decety & Keenan, 2006, p. 1)

Decety and Keenan (2006) argue that the terms 'social' and 'neuroscience' have been traditionally defined in similarly broad and liberal ways. First, they argue that "behaviors and cognitions studied under the umbrella 'social' are diverse" and that "social research is an expansive, diverse, and complex domain” (2006, p. 1). This is reasonable given that humans are highly social animals and thus many of their behaviours are motivated in one way or another by social processes. In fact, for most of our evolutionary history, humans have evolved and lived under circumstances in which they were almost never alone, and thus constantly behaved in a social way (this, however, is not to say that all human behaviour is social). Indeed, humans have adapted to have an extraordinary sensitivity in detecting social exclusion from their group, which is said to lower probability of survival (Williams, 2007). I will elaborate on this topic in chapter 2. Returning to the definition of sociality (and social cognition), it might be argued that because social processes are so ubiquitous, diverse, complex, and therefore hard to define, sociality is better conceptualized as a Roschian prototype concept (Rosch & Mervis, 1975), which has fuzzy boundaries and is not, at this point at least, definable. At the same time researchers have a good idea of what they mean by 'sociality', even if they cannot define it in terms of classic Aristotelian logic.

What constitutes ‘neuroscience research’ - the scientific study of the nervous system- also needs to be defined. Decety and Keenan (2006) rightly argue that neuroscience should be defined in a way that includes indirect measures of neurological underpinnings of behaviours. They argue, for example, that the study of animal behaviour can be defined as neuroscientific research if it adds knowledge to the workings of the nervous system.

It is however possible (and probably easier) to distinguish between behaviours such as feeding, which is clearly not necessarily socially motivated, and behaviours such as smiling at someone, which is clearly socially-motivated. One could, perhaps, question whether all phenomena that are investigated under the label of social neuroscientific research are necessarily social, or whether similar behaviour could also be found in other species that are not (as) social. For example, one topic in social neuroscience that is not necessarily social is that of biological motion detection (Cacioppo & Berntson, 2004). Humans are very good at inferring information from so-called “point-light walkers” that display biological motion, but that do not give any static form cues (Grossman & Blake, 2002). Viewing point-light displays activates brain areas that are associated with social functions such as viewing faces (Grossman & Blake, 2002). However, the function of being good at detecting biological motion might not be aimed at being social; i.e. humans might not need biological motion detection to be social, but rather for survival (to enable rapid detection of predators). In contrast, some researchers propose that biological motion detection goes beyond survival. Puce and Perrett write:

> Primates, being social animals, continually observe one another’s behaviour so as to be able to integrate effectively within their social living structure. At a non-social level, successful predator evasion also necessitates being able to 'read' the actions of other species in one's vicinity. The ability to interpret the motion and action of others in human primates goes beyond basic survival […] (Puce & Perrett, 2003, p. 435)

However, it might not necessarily be the case that detecting biological motion goes beyond basic survival. Many animals that detect biological motion are traditionally seen as non-social, especially those belonging to non-mammalian species (although even these species sometimes exhibit behaviour that could be construed as social: for instance, in a social setting aplysia, a species of sea slug, stops responding to palatable food that is too tough to swallow, but when the same aplysia are socially isolated this learning is blocked – which has been shown to be caused the absence of pheromones secreted by other aplysia (Cacioppo, 2002)). Sometimes these non-social species have hard-wired biological motion detection for other animals that are prey as well as predators: frogs, for example, have neurons that code for the movements of typical prey and predators, and these neurons trigger behavioural programs that make them approach or avoid (see for instance, Zupanc, 2004). To call this a social response might go against our intuitive definition of what 'social' means, as those responses are likely useful for the survival of solitary animals. Moreover, humans can infer properties such as gender and heaviness from point-light walkers and are also able to infer emotional states and even personality traits (e.g. relaxed/nervous, happy/sad) (Brooks et al., 2008). However, one might speculate that the inferences of emotional states are drawn from culture-specific stereotypes, and only information that is very important for estimating threat - gender and heaviness - can be reliably inferred independent of culture. In fact, some argue that point-light walker displays contain a wealth of information (e.g. emotions) that humans are very inefficient at exploiting (Gold, Tadin, Cook, & Blake, 2008). Moreover, there are some biases that probably relate directly to survival: for instance, it has been shown that there is a bias towards perceiving that a male point-light walker is approaching the viewer, whereas a female point-light walker is more readily interpreted as moving away from the observer (Brooks et al., 2008). Taken together, these findings might suggest that biological motion detection is necessary for survival, also because humans have constituted one of the main threats for the survival of other humans (Fry & Söderberg, 2013). However, this perceptual ability has likely only in more recent evolutionary history become important for sustaining social relationships, so that the original evolutionary purpose was more general agency detection and the usefulness for sociality is only 'piggybacking' on this ability.

In response, perhaps, to lines of research such as biological motion being seen as social neuroscience, it has been suggested that social neuroscience research should be more firmly based in social psychology (Todorov et al., 2006). It becomes more evident why this call is needed if one looks at other research areas of social neuroscience, where many studies are focused on low-level, automatic, spontaneous processes that occur during the first seconds of a social interaction. For instance, according to Cacioppo and Bernston's *Social Neuroscience: Key readings* (2004), it would appear that topics in social neuroscience include: face processing, biological movement, and processing of facial signals. While this research is important, only a few topics like theory of mind, causality and animacy reasoning, or cognitive appraisal, deal with higher order, knowledge driven top-down social processes. Indeed, Todorov, Harris, and Fiske (2006) argue that the most important question for social neuroscience is how bottom-up and top-down processes interact. It is also interesting to note here that that theory of mind research comes from developmental psychology, and some researchers call attention to the fact that, despite the similarities between this field and social psychology's attribution research, there has not been much dialogue between those two branches (Malle et al. as cited in Todorov et al., 2006). Todorov et al (2006) highlight that much of social psychology is about the power of the situation over behaviour, and calls for creating realistic social situations in social neuroscience experiments. For instance, they report studies showing that meeting a confederate before a brain scan can influence later brain activity; other studies have successfully deceived participants into thinking that they were interacting with other participants in other brain scanners. They assert that the future "belongs to fMRI experiments able to study real social interaction in real time" and that "behavioural research on person perception has been moving toward assigning a smaller and smaller role to deliberate, computationally taxing strategies until now, the future seems to belong to the goal dependence of social cognitive processes, including relatively automatic processes" (Todorov et al., 2006, p. 78). Other researchers have recently highlighted the fact that humans have different responses to encountering people in real life and seeing a life-size photo of people and that the latter is more like the experimental setup of most fMRI experiments (Risko, Laidlaw, Freeth, Foulsham, & Kingstone, 2012).

This trend of focusing on more automatic processes in social neuroscience can even be seen among the most basic social behaviours social neuroscience is concerned with. For example, there has a lot of research into the human mirror neuron system, a proposed class of neurons. These single neurons fire during both execution and observation of actions (Rizzolatti, Fogassi, & Gallese, 2001). One interpretation of this finding is that since we automatically represent other people's movement (including facial expressions) whenever we see them, this simulation underlies social cognition processes such as theory of mind, imitation and empathy (Iacoboni & Dapretto, 2006; Rizzolatti et al., 2001).

However, the existence and significance of mirror neurons in humans was controversial (Dinstein, Gardner, Jazayeri, & Heeger, 2008), but see (Kilner, Neal, Weiskopf, Friston, & Frith, 2009). More recently, recording extracellular activity in human patients has shown that there are indeed neurons that respond to action execution and action observation (Mukamel, Ekstrom, Kaplan, Iacoboni, & Fried, 2010), corroborating previous evidence from single cell recordings in monkeys (di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992).

At the same time, however, there has been relatively little research into the neural basis of similar behaviours that explain more complex phenomena like cooperative movements and joint actions, that is, those movements during which people do not mirror, but where one person intuitively performs a movement that opposes that of another person, as might happen during a dance (see for instance Sebanz, Bekkering, & Knoblich, 2006). This is not to say that any research into automatic, bottom-up behaviours and the like is not important, but that social neuroscience should maybe also entail more complex social behaviours based on realistic social settings as proposed by Todorov et al (2006). In this thesis, I aim to integrate findings from experiments that examine both high-level and low level processes necessary for social interactions.

Considering that many behaviours are social and that every social behaviour must be triggered by the nervous system, the field of social neuroscience, in theory, should deal with and explain many behaviours. This is important to consider when asking how *social* social neuroscience is, because much behavioural research then becomes a type of social neuroscience investigation. These research activities may vary in sociality, which together with the fuzzy definition of social behaviour, makes it even harder to say whether something is social neuroscience. For instance, is anything that involves interacting with another agent automatically "social", such as one-time encounters in which certain facial expressions signal the motivation to aggress or is “social” something long-term, about living and cooperating together? It is difficult to answer that question.

Moreover, it is a good idea that scientists in the field define ‘social’ and ‘neuroscience’ very liberally and inclusively, to allow for interdisciplinary collaborations and contributions from researchers with different backgrounds and methodological approaches. However, as I have suggested here, with the example of biological motion detection research, not all fields that are labelled as social neuroscience are necessarily social at all. Moreover, some researchers argue that social neuroscience should be more firmly based on social psychological research, as there is a trend towards only looking at very basic, low-level, bottom-up behaviours, while neglecting more complex deliberate behaviours. Indeed, in his review of the social brain, Frith (2007) similarly argues that various components of the social brain are not specifically social (apart from the mirror neuron system).

1.1.2. Overview of the thesis

This thesis aims to contribute to the field of social neuroscience by taking into account the above mentioned considerations. I have performed studies that analyse complex deliberate behaviours (Chapter 2) to more intermediate behaviours (Chapter 3-4), and more automatic processes (Chapter 5). I will attempt to explain all of these processes using a hierarchical predictive coding framework (Chapter 6). Some more assumptions and theoretical considerations that went into the study designs in this thesis will be discussed in the next section, as they pave the way for the experimental chapters (Chapter 2-5). Specifically, I will address the need for increased emphasis on functional integration (i.e. connectivity research) when analysing neuroimaging data and also the need for creating ecologically valid research paradigms in social cognition, which can nevertheless be embedded in domain-general theories.

1.2. Functional specialisation of social cognition

Much recent research in (social) cognitive neuroscience has concerned the functional localization or functional mapping (see Friston, 2003) of brain structures that process (social) information (Passingham, Rowe, & Sakai, 2013). Functional specialization describes the assumption that there are certain brain areas that are specialized for certain brain functions (Friston, 2002b). Relatedly, functional localization is the attempt to characterise the functional specialization of different brain areas. The ‘social brain’ comprises a network of regions that are involved in social processing (Frith, 2007). Many reviews in social neuroscience identify areas that are involved in social perception, cognition and processing (Amodio and Frith, 2006; Frith, 2007; Moll et al., 2005; Saxe, 2006b, Mar, 2010, Van Overwalle, 2009). Among the areas most often reported in studies on mentalizing, are the medial prefrontal cortex (mPFC), the temporoparietal junction (TPJ), the superior temporal sulcus (STS), and the temporal poles (Amodio and Frith, 2006). However, many different tasks from fields other than social cognition have been shown to activate these areas reliably (Amodio and Frith, 2010, Egner, 2011). Moreover, quantitative meta-analyses of activations from social neuroscience studies show that there might be other cortical areas that are reliably activated during mentalizing tasks in addition to the previously mentioned areas, namely bilateral angular gyri, posterior cingulate cortex, precuneus, and the left inferior frontal gyrus (IFG) (Mar, 2011). Another recent activation likelihood estimation (ALE) meta-analysis concludes that even the cerebellum is critically implicated in fMRI social cognition tasks (Van Overwalle, Baetens, Mariën, & Vandekerckhove, 2013). This raises doubts over the predominant idea that higher-level cortical areas are uniquely implicated in social cognition fMRI studies and hints at more functional integration of higher-level functions than perhaps previously assumed. Other social tasks such as biological motion processing, face processing, and emotional processing involve additional cortical and subcortical areas. In the following section, I will discuss the nature of these tasks.

1.3. Explaining ecologically valid tasks in domain-general terms

There is typically a trade-off in cognitive neuroscience between research with high generalizability and high ecological validity. One can examine brain activations associated with very basic but highly generalizable processes (e.g. by creating computational models like reinforcement learning models, Behrens, 2009). These typically result in robust and reliable activations. Highly ecologically valid tasks examine brain activations that occur when participants are engaged in naturalistic tasks that resemble the demands of the processes humans are confronted with in everyday live, and these processes supervene on the more basic processes

(Schilbach, 2010). These processes might also have more clinical relevance than the more generalizable tasks. However, most tasks in social neuroscience are artificial in the sense that humans would not encounter them in the environment that they have evolved in nor in their daily life. I argue here that, when one designs ecologically valid tasks that resemble naturalistic settings, and place constraints on participants that they would encounter in the natural environment, care must be taken in the interpretation of the findings. Specifically, it is important to take a more domain-general perspective when integrating the findings of the study with an explanatory framework for the functional role of the brain region with respect to the social cognitive process being studied. For instance, previous research has depicted the mPFC as a special module for social cognition (see Amodio and Frith, 2010 for a review), but mPFC activations have been associated with many paradigms from a wide variety of fields (Amodio and Frith, 2010, Egner, 2011). To integrate these findings, domain-general theories of the mPFC have been proposed. Amodio and Frith hypothesize that the mPFC represents information about goals and behaviours in the context of complex external contingencies that can be, but are not necessarily, social goals (Amodio and Frith, 2006).

Complex ecologically valid social tasks can be broken down into several more basic components. For example, a task in which participants are asked to think about themselves vs. others (Jenkins, Macrae, & Mitchell, 2008) or think about a certain emotion vs. some other emotion (Burnett, Bird, Moll, Frith, & Blakemore, 2009) will have some confounds that cannot be controlled for, such as the effort involved in thinking about a certain emotion or about other people. Ward (2012) argues that caution is needed when interpreting differences in BOLD signal in social neuroscience fMRI studies as they might reflect task difficulty, attention or strategy to solve task (Ward, 2012). Consider a complex ecologically valid social task that consistently shows e.g. mPFC activation. If one were to deconstruct such a complex task into its subcomponents, only a subset of those components that make up this task will be directly related for the social cognitive process of interest, rather than the entire set of basic components that make up the ecologically valid task. Thus, there might be superfluous components or processes that are not necessarily special to a social cognitive process, such as task difficulty, that could drive activations.

A comprehensive review of the computational literature (Egner, 2011), shows that the more basic processes in which the mPFC is involved are: error detection; conflict monitoring and the prediction of error likelihood, as well as a variety of computations underlying reinforcement-guided decision making, such as error-driven negative reinforcement learning signals; the calculation of action values on the basis of previous action-outcome associations and the amount of effort involved in the action, action value prediction error; and the estimation of environmental volatility that determines the rate of action value learning (Egner, 2011). It has been proposed that a single computational description that unifies all these diverse processes is negative surprise i.e. when actions do not produce the expected outcome (Egner, 2011). Most tasks in social neuroscience are only conceptually matched tasks, like thinking about the self vs. others, and thus they might be driven by more general differences that have been demonstrated to activate mPFC, such as effort. In other words, activity in a particular area might not be because this is a special social cognitive module, but rather that some confounding computation is likely to co-occur with social tasks (Behrens, 2009). This is especially a problem when the tasks are not ecologically valid, like in social neuroscience tasks in which participants read sentences or passively view static faces, and the constraints of the experimental setting make one task condition more effortful, surprising or interesting than another condition. In other words, cognitive neuroscience experiments often show participants sentences or other non ecologically valid stimuli in a scanner that prompt them to think e.g. about themselves vs. others. The results of some not very controlled experiments can be confounded when one condition is harder to parse (there are few stimulus controls in social cognitive neuroscience experiments) and this might drive activity in mPFC (which in other areas of cognitive neuroscience are associated with effort). Thus, one cannot conclude that generally, when people think e.g. about themselves vs. others in their daily life, that they will have more activity in the mPFC.

Nevertheless, even though brain activity during social cognitive tasks can in principle always be explained with more domain-general theories that have higher generalizability (Behrens, 2009), results of high-level social tasks are informative and needed. Not only do they assist in the creation of more constrained models (Behrens, 2009), but they also reveal which brain regions are involved during ecologically valid settings. This is important to know since brain structures have evolved in response to evolutionary challenges that are naturally occurring. A more domain-general theory of the fusiform face area (FFA) proposes that it is particularly responsive to high-level visual expertise and this explains why it responds both when most participants view faces (Kanwisher, 1997), but also when bird-watching experts view birds or car-experts view cars (Gauthier, 2000). The reduced domain-general theory is simpler and explains more data. However, the more ecologically valid theory is nevertheless informative, as humans are experts in face recognition and the FFA has probably primarily evolved to meet the adaptive challenge of face recognition. This insight gave rise to the more domain-general model of the FFA as an area implicated in visual expertise. For this reason, it is desirable to move one's high-level experiments towards high ecological validity in order to be predictive of real life (social) situations (Przyrembel, Smallwood, Pauen, & Singer, 2012) and have translational value such as clinical usefulness. However, when investigating social processes and localising brain structures responding to the demands of ecologically-valid social tasks, it is still important to find domain-general theories that fit the response profile to guide future research. Finally, as Ward argues (Ward, 2012), there is currently still no consensus view on whether there are domain specific neural mechanisms of social cognition. It might be that general purpose neural systems are responsible for social cognition and there are no areas of the brain that are uniquely social (Rushworth, Mars, & Sallet, n.d.).

As mentioned above, most studies in social neuroscience have traditionally used paradigms that are quite removed from real social interactions. In this thesis, I aim to use more ecologically valid paradigms in order to explain real life social behaviour, but also to interpret the activations in terms of domain-general theories.

1.4. Functional integration of social cognition

After localizing brain regions associated with social cognitive tasks, the next step in order to gain a more complete understanding of the social brain is to look at the functional integration of different brain activations (Friston, 2002a). No neuronal subpopulation operates independently from other parts of the brain and social processing - just like any other cognitive processing - is likely spatially distributed as well as temporally dispersed. In order to understand the neural mechanisms of social cognition one needs to test models that make explicit references to the inputs that innervate particular regions, and the other areas to which these regions project (Adolphs, 2010). More concretely, it is useful to investigate whether activations in higher-level areas are modulated by activity in primary sensory areas and vice versa. Functional integration can be elucidated by investigating structural, functional and effective connectivity between brain regions. Structural connectivity research investigates which areas are anatomically connected to what other brain areas. Functional connectivity is often defined as the description of statistical dependencies of (neural) activity in spatially distinct brain areas to infer connectivity between them (Stephan & Friston, 2010). This is often done by investigating how far activity in different areas is temporally correlated in resting state fMRI, without an experimental manipulation (Fox et al., 2005). In contrast, effective connectivity is defined as connectivity that is based on a mechanistic (causal) model, that describes the influences of one area over another (Stephan & Friston, 2010), which requires an experimental manipulation. This requires the model to be the simplest possible circuit diagram that is dependent on time and experimental perturbations (Stephan & Friston, 2010).

There are principally two different ways of analysing task-dependent (effective) connectivity: non-directional and directional connectivity. A classic example of non-directional effective connectivity is psycho-physiological interaction (PPI) analysis (Friston et al., 1997). PPI analysis can supplement traditional statistical parametric mapping analysis to investigate task-dependent changes in connectivity between brain areas in different psychological contexts. PPI is based on linear regressions. A psychophysiological interaction takes please if the physiological source (i.e. the brain activity in a certain brain area) can predict brain activity in another brain area better within a certain psychological context than in another (i.e. a certain experimental condition). This suggests that the first area exerts influence on the other brain area in a task dependent fashion, as opposed to merely being suggesting correlated brain activity.

In contrast to non-directional connectivity as assessed by resting state fMRI, effective connectivity methods such as dynamic causal modelling (DCM) (Friston, Harrison, & Penny, 2003) provide additional information about the directionality of the influence of one area over another. DCM estimates the experimental modulation of (intrinsic) self-connections or (extrinsic) forward and backwards connections between brain regions that are active during a particular task in a directional manner. One can then infer whether experimental manipulations affect top-down or bottom-up influences, or both.

The generative model used by DCM is based on coupled bilinear differential state equations modelling distributed brain activity and canonical haemodynamics for each region (Friston, Harrison, & Penny, 2003). The change in the state vector *x* in time can be modelled with the following equation, which has *u* as an experimental input (Stephan et al., 2007):

$$\frac{dx}{dt} = \left(A + \sum_{j=1}^{m} u_j B^{(j)} \right) x + Cu$$

Here, the three parameters are estimated in classical bilinear DCM. The first parameters (the DCM.A matrix) estimate fixed connections between brain regions, i.e. the effect that one brain region has upon another, in a baseline condition (this sometimes referred to as context-independent, average, or intrinsic connectivity). The second parameter (DCM.B) estimates the modulation of the fixed connections between brain regions as a result of a particular task condition, i.e. the impact of the task on the connectivity between brain regions, rather than the effect that the task has on specific brain regions. Finally, *u* denotes the inputs to the model and the third parameter (DCM.C) indexes the direct influences on an area. The extrinsic driving input usually consist of a sensory contrast that sets the system in motion, whereas the modulatory contrast are of a more attentional nature and then change the coupling (Stephan et al., 2010). A non-mathematical metaphor for this model would be two chambers (here brain areas) separated by a gate, that are flooded with water (here brain activity), where water is always entering the first chamber (driving activity is often entering a sensory region) and then only when a particular gate is opened, that means when an experimentally induced attentional modulation occurs, water enters the next chamber (activity enters a region higher in cortical hierarchy). In the chapters 4 and 5, we will use this method to investigate effective connectivity.

1.5. Paradigms to study social behaviour

I will now describe some more general background regarding the paradigms employed in the thesis. First, I will review behavioural game theory with a focus on games that capture aspects of social interactions like trust, then I will review theory of mind tasks with a focus on perspective taking paradigms and how performance on these tasks changes throughout development and the accompanying neural changes. Finally I will introduce the Bayesian hierarchical predictive coding framework with which I will attempt to integrate the findings in this thesis.

1.5.1. Game theory and economic games

The field of behavioural game theory (Camerer, 2003) and most recently (social) neuroeconomics (Fehr & Camerer, 2007), which uses the mathematical formalism of game theory to quantify social behaviour and relate it to measures of neural activity, has gained increasing popularity as a way of studying (human) social behaviour. Game theory lends itself well to the study of social behaviour because variables that are socially meaningful (e.g. trust) can be formalized mathematically. Moreover, monetary transactions endow the experiment with ecological validity due to participants being particularly motivated by financial rewards.

The most classic paradigm of behavioural game theory is the prisoner's dilemma (Axelrod & Hamilton, 1981), which formalizes the problem of cooperation first characterized in behavioural ecology (Camerer, 2003). In the prisoner's dilemma, two (or more) players (or prisoners) have to decide whether they want to cooperate with the other person or defect. Crucially, even though it might not be immediately apparent, these economic games often capture real world challenges that humans have faced (and still face) in the environment they have evolved in. This is however not to say that all economic games as they are played always accurately reflect behaviour outside of the laboratory. For instance, recent work has shown that in natural field experiments there is no altruistic giving in dictator games (Winking & Mizer, 2013), which is different from experimental findings in the lab. The prisoner's dilemma, in contrast, mirrors real-world situations such as that of two people cooperating during during persistence hunting, a uniquely human hunting technique, where a group of humans pursue an animal to the point of exhaustion (Liebenberg, 2008). The hunters can decide to cooperate and thereby achieve a bigger payoff through division of labour and the capture of a bigger animals would be caught than when acting alone, but then they run the risk of being exploited in the process. Alternatively, they can choose not to cooperate, which is equivalent to working by themselves, and achieve a smaller payoff (in this example a smaller animal) without running danger of being exploited by so-called free riders. Alternatively, they can 'defect' and exploit the partner's trust when the partner is cooperating and free ride (Axelrod & Hamilton, 1981). Due to humans' highly cooperative nature, they play versions of these economic games frequently in every day contexts. Moreover, such economic exchanges often require participants to take the perspective of the person they interact with in order to cooperate successfully. For instance, in the trust game (chapter 2), in order to know whether to trust someone, one has to reason about the mental states of another person, about whether the other person is likely to cooperate or defect. This reasoning might not always rely exclusively on 'cold cognition', but can be emotionally informed- as in nature most social interactions are iterated (repeated) games. Repaid trust by cooperation from the interactant can lead to emotional reinforcement of cooperative behaviour with a particular interactant (King-Casas et al., 2005). These trust processes can then become automatic without having to explicitly engage in theory of mind processes to figure out whether someone is trustworthy. By using game theory and experimental behavioural economics one can study a myriad of social phenomena including altruism, trust, cooperation, or the utility of social interactions.

In social neuroeconomics, dependent variables from game theory can more conveniently be parametrically modulated, because they can be measured in terms of monetary values, and then correlated with neural measures (see for instance King-Casas et al., 2005). This has been shown to be a successful research paradigm and many studies have capitalized on the fact that computational models elicit robust neural responses during observational learning that are similar to activation elicited by decision-making paradigms in reward-related areas (King-Casas et al., 2005; Rangel, Camerer, & Montague, 2008). Formal computational models have not only been applied to social observational-reward-learning (i.e. predicting rewards from other people), but also to model action tendencies of others and inferring the hidden (intentional) mental states or traits of others (Dunne & O'Doherty, 2013).

1.5.2. Theory of mind tasks

Theory of mind abilities were first studied in primates (Premack & Woodruff, 1978), and subsequent theoretical work proposed that special tasks are needed to study theory of mind (Dennett, 1981). The earliest theory of mind tasks that were used were called false belief tasks (Baron-Cohen, Leslie, & Frith, 1985) and researchers have been interested in the neural correlates of false belief tasks since the early days of neuroimaging using PET (Fletcher et al., 1995; Happé et al., 1996). In a typical false belief task (Baron-Cohen et al., 1985), participants read a story involving different characters that either have different beliefs about the state of the world or have mental states that are divergent from those of the participant. Participants are asked to indicate what the mental state of the character in the story is. For instance, the prototypical false belief task is the Sally-Anne task in which Sally puts a marble in a basket while Anne is watching, and Sally then leaves the room. Unbeknownst to Sally, Anne takes the marble out of the basket and puts into a cupboard. In the story, Sally, then re-enters the room. The participant is asked where Sally will look for the marble. The answer is that she will look in the basket as this is where she left the marble and expects it to be. Many children below the age of four years typically perform poorly on these tasks and cannot state another person's belief correctly if it is divergent from their own mental state (Wellman, Cross, & Watson, 2001). Similar results were shown in children with autism who are older than five (Baron-Cohen et al., 1985; Yirmiya, Erel, Shaked, & Solomonica-Levi, 1998). However, there is substantial evidence that children who do not pass false belief tasks do actually have a theory of mind. For instance, evidence from nonverbal tasks suggests that 15-month-old infants are able to predict an actor's behaviour on the basis of her true or false belief (Onishi & Baillargeon, 2005). Adults have been shown to calculate other people's spatial perspectives automatically (Samson et al., 2010). Infants have also been shown to calculate what another person can or cannot see (Flavell, Abrahams Everett, Croft, & Flavell, 1981; Moll & Tomasello, 2006) as have primates (Hare, Call, & Tomasello, 2001). Therefore, the ability to infer mental states and the explicit reporting of those states appears to develop in a step-wise fashion during the first five years of life (Frith & Frith, 2003), after which children are able to reason explicitly about theory of mind.

However, this ability does not always lead to automatic, online use of theory of mind information even in adolescents and adults (Apperly, 2012; Apperly et al., 2010a). For instance, Keysar and colleagues (Keysar, Barr, Balin, & Brauner, 2000; Keysar, Lin, & Barr, 2003) designed a task in which participants were faced with a set of shelves containing objects that were either visible or not visible from the viewpoint of a "Director" (a confederate). The director asked participants to move objects in the shelves and critical instructions required participants to use information about the director's viewpoint to correctly interpret their instructions. In this Director task, adults did not reliably use their theory of mind knowledge to interpret the intentions of others (Keysar et al., 2000, 2003). Surprisingly, around 50% of the time, they failed to use information about the director's perspective and instead erroneously used their own (egocentric) viewpoint only, when trying to follow the director's instruction. These results were replicated using a computerised version of the paradigm and controlling for the inhibitory control demands of the task with a matched No-director condition (Apperly et al., 2010b). Furthermore, egocentric errors in this task are accentuated in adolescence, with an even stronger egocentric bias observed in 14-17 year-olds than in young adults (Dumontheil, Apperly, & Blakemore, 2010). These results suggest that the ability to take another person's perspective in order to select appropriate actions is still undergoing development at this relatively late stage. The Director task differs from other theory of mind tasks in that it requires participants to have a functioning theory of mind, to compute the perspective and intentions of another person (the director), and use this theory of mind information in concert with other cognitive processes such as executive functions to overcome their egocentric bias and select the appropriate response quickly and accurately (Apperly et al., 2010b). It is proposed that it is this interaction between theory of mind and executive functions that continues to develop in late adolescence, and is still prone to errors in adults (see Dumontheil et al., 2010). All this suggests that humans often perform relatively poorly on theory of mind tasks. Surprisingly, however, there is also some evidence that preverbal infants (Onishi & Baillargeon, 2005) can understand false beliefs. Moreover, even chimpanzees have some theory of mind abilities (Call & Tomasello, 2008). Finally, people seem to rapidly and involuntary compute what other people see in the sense that their reactions and reaction times are biased (Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010). Given this evidence for low-level theory of mind, then why do children, people with autism, fail theory of mind tasks and adolescents and even adults often fail to spontaneously pass such perspective taking tasks? There have been attempts to explain poor performance of people with autism in theory of mind tasks in terms of motivational states, and some evidence suggests that children with autism actually do track other people's beliefs but only in competitive games (Peterson, Slaughter, Peterson, & Premack, 2013). Another reason for failure on theory of mind tasks, which we examine in Chapter 3, might be that the mental state of another entity is computed, but then one's own perspective is too salient or prepotent to be inhibited (Apperly et al., 2010a). There have been only few attempts to disentangle these processes of computation of another entity's mental state and the selection of the correct perspective and we will focus on these in chapter 3 and 4. The development of theory of mind throughout adolescence is accompanied by neural development, which will be examined in the next section.

1.5.3. Brain development in adolescence

While many studies over the past 30 years have investigated the development of mentalising in infancy and childhood, pointing to step-wise changes in social cognitive abilities during the first five years of life (Frith & Frith, 2007), only recently have experimental studies focused on development of the social brain beyond early childhood. Recent cognitive neuroscience studies have focused on adolescence as a period of profound social cognitive change. Adolescence is defined as the period of life between puberty and the attainment of a stable, independent role in society (Steinberg, 2010). The recent cognitive neuroscience studies support evidence from social psychology that adolescence represents a period of significant social development. Adolescence is characterised by psychological changes in terms of identity, self-consciousness and relationships with others (Steinberg, 2010). Compared with children, adolescents are more sociable, form more complex and hierarchical peer relationships and are more sensitive to acceptance and rejection by peers (Steinberg & Morris, 2001). Together with the hormonal changes, there are often changes in the (social) environment during adolescence, such as school changes or one's peer group becoming more important relative to one's family, and these different hormonal and situational influences are likely to influence each other. While the causes of these social changes in adolescence are likely to be multi-factorial, development of the (social) brain is likely to play a significant role.

Adolescence is a period of social and psychological development during which social awareness and behaviour undergo profound change (Brown, 2004; Eisenberg & Morris, 2004). At the same time higher cognitive control and reasoning abilities mature (Crone & Ridderinkhof, 2010; Crone, 2009; Dumontheil & Blakemore, n.d.; Luna, Padmanabhan, & O'Hearn, 2010). As well as alterations in hormone levels and social environment, a more proximal cause of these developmental changes are structural changes taking place in brain areas involved in social cognition, including the medial prefrontal cortex (mPFC), the superior temporal cortex, and the temporo-parietal junction (TPJ) (Frith & Frith, 2003; Gallagher & Frith, 2003; Saxe, 2006), and in brain regions involved in cognitive control tasks, in particular lateral parts of the PFC. Notably, frontal and temporal lobes undergo protracted structural development in humans (Giedd et al., 1999; Gogtay et al., 2004; Mills, Lalonde, Clasen, Giedd, & Blakemore, 2012; Shaw et al., 2008; Sowell et al., 1999). Developmental functional imaging studies of mental state attribution have consistently shown that mPFC activity during a variety of mentalising tasks (e.g. understanding irony or thinking about one's intentions) decreases between adolescence and adulthood (Blakemore & Robbins, 2012; Blakemore, 2008; see Burnett, Sebastian, Cohen Kadosh, & Blakemore, 2011 for reviews). In addition, there is evidence of developmental changes in functional connectivity between mPFC and other parts of the mentalising network during adolescence (Burnett & Blakemore, 2009). The age-related changes lie in the part of the mPFC superior to z=0 labelled anterior rostral medial frontal cortex (MFC), which is recruited by tasks involving self-knowledge, person perception, and mentalising, in contrast to a more ventral 'orbital MFC' region and a more posterior 'posterior rostral MFC' region (Amodio & Frith, 2006). Other groups use slightly different subdivisions of the mPFC, with a more superior border at z = 20 (Van Overwalle, 2009) to distinguish between "ventral mPFC" and "dorsal mPFC". This dissociation has been related to a distinction between mentalising judgments made towards the self or similar others vs. dissimilar others (Mitchell, Macrae & Banaji, 2006; Jenkins, Macrae & Mitchell, 2008; Tamir & Mitchell, 2010; Van Overwalle, 2009).

Until recently there had been little research on the development of structural, functional and effective connectivity in adolescence, even though this has potential to elucidate differences in behaviour, cognition and mental state. Functional connectivity analysis examines the statistical dependencies between different brain areas. A baseline measure of this, in the absence of any task, is referred to as resting state functional connectivity (rsfMRI). One recent study showed that the connectivity between nodes of the default mode network (the network of brain regions that are active during baseline) changes during adolescence (Jolles, van Buchem, Crone, & Rombouts, 2011; Lopez-Larson et al. 2011; also see Fair et al. 2008). Another rsfMRI study found that the influences between different networks reduce with age and that this might reflect enhanced within-network connectivity and more “efficient” (and thus reduced) between-network influences (Stevens et al. 2009). There are surprisingly few developmental task-related neuroimaging studies that have employed effective connectivity methods. Only one previous social brain study has investigated effective connectivity development over adolescence (Burnett & Blakemore, 2009). This study showed that functional connectivity within the mentalising system was higher during social versus basic emotions in adults and adolescents, and that there was a developmental reduction in functional connectivity within the mentalising network, possibly due to increased specialisation and increasingly “efficient” between-network influences (cf Stevens, Pearlson, & Calhoun, 2009). Two recent studies using effective connectivity methods found that, compared to adults, both children (Bitan et al., 2006) and adolescents (Hwang, Velanova, & Luna, 2010) show weaker top–down modulatory influences from frontal areas.

1.6. Bayesian Hierarchical predictive coding approach to explanation of the findings

Google Inc. recently announced the implementation of a new image format called WebP, which uses predictive coding to decrease the file sizes of average images and videos by about 30 percent (Ginesu, Pintus, & Giusto, 2012). Interestingly, this will reduce the cost of global internet traffic significantly due to video streams having a significantly lower size. How does this compression algorithm work? A simplified explanation is that the algorithm makes predictions about the hidden states of the world based on an internal model of it. This model predicts that there will be consistency from one pixel to another, so that it expects that if one pixel is black the neighbouring pixel is likely also black. If this prediction is true, it mostly needs to save are values of zeros, which are very efficient to compress. If the prediction is false, then there will be a prediction error. Because the file only saves the prediction errors and not the actual colour of the pixel, the file size grows more rapidly the more prediction errors there are. Put simply, if the next pixel is dark grey it saves a value of e.g. -1 (a small prediction error that, in theory, needs 1 bit to save) and if the next pixel is blue is it saves a value of say -100 (a large prediction error, which needs approximately a byte to save). Because photographs and videos of the world contain consistent patterns (e.g. a continuously blue sky), this procedure is computationally efficient. The more complex, random, or unpredictable an image (cf. entropy), the more prediction error there will be and the more disk space it uses. Recently, it has been proposed that the brain also uses Bayesian (hierarchical) predictive coding to efficiently process incoming sensory data (Brown & Brüne, 2012; Clark, 2012; Friston, 2010). Perhaps the most compelling argument that the brain uses predictive coding is that this preprocessing increases the metabolic efficiency of the brain. If the data were unfiltered, the information hitting the retina would be 36 Gb/s, preprocessed it is approximately 20 Mb/s of useful data (Sengupta, Stemmler, & Friston, 2013). Thus, the preprocessing of incoming sensory data would reduce the metabolic cost by a factor of about 1,500 (Sengupta et al., 2013).

To stick with the picture/vision example, consider the visual illusion shown in figure 2. The background is a gradient and the bar in the middle seems to be a gradient too. However, in actuality, the bar in the middle is uniform grey, which one can see if one takes the context away and superimposes the background with a uniform colour. Without going into more detail about the visual illusion, this suggests that colour is interpreted based on the context-sensitive cues and that the brain makes certain assumptions based on an internal model of the world. These illusionary percepts have been argued to be Bayes optimal i.e. what we perceive is not actually a failure of perception but instead the best possible explanation of the incoming sensory data (Brown & Friston, 2012). We owe these models to our evolutionary past and influences that trigger learning during development. In other words, if recognizing that the middle bar is isoluminant would have been selected for in our ancestral environment or if we would have been exposed to such illusions from a very early age and recognizing the isoluminosity would have been positively reinforced, then the internal model probably would have updated and we would not have such strong illusions. Some of our models are more plastic and the predictions of the model are update frequently (cf. Kalman filtering). For example, if my model of a cup is such that I believe it to be full, when in reality, the cup is empty – I will apply too much force when picking up the cup. I will update my model of the world based on this prediction error and next time I pick up the cup, I will apply less force, which will result in less prediction error.

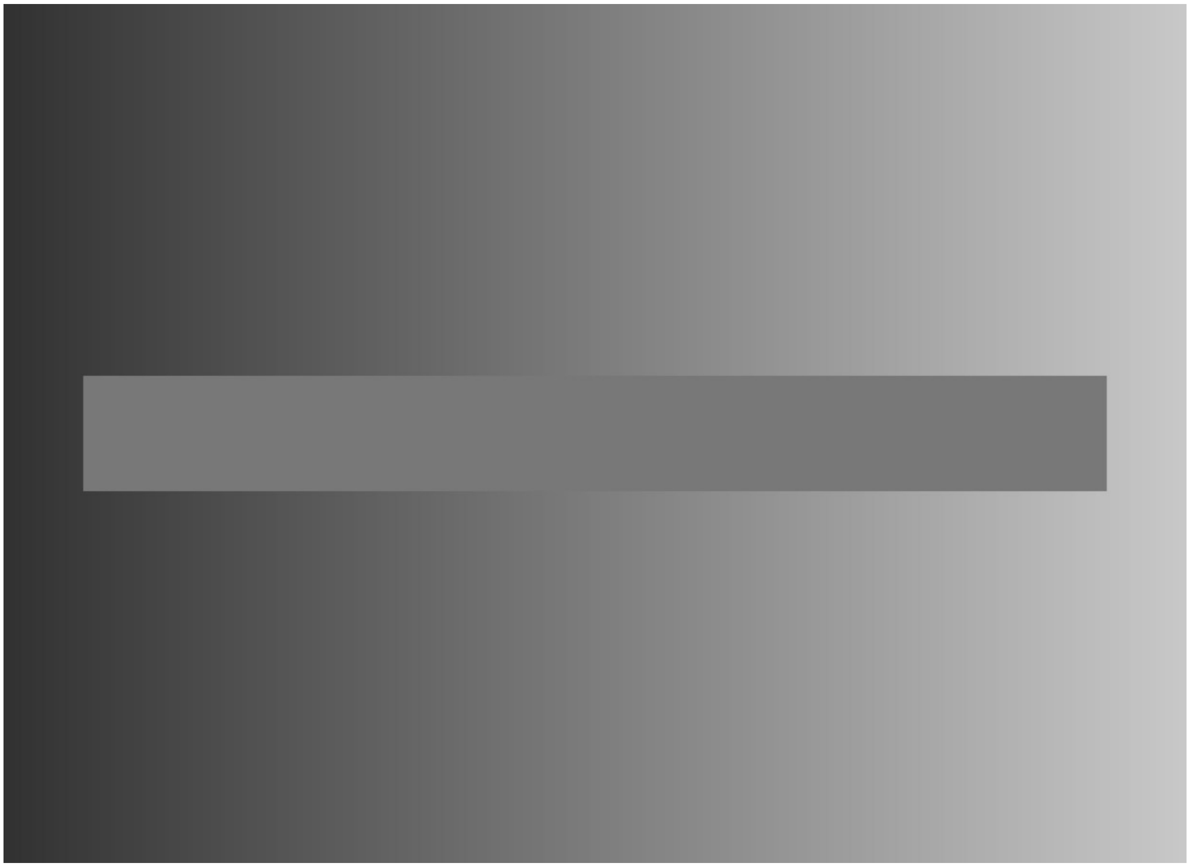


Figure 2: Simultaneous Contrast Illusion. Here the background is a colour gradient and progresses from dark grey to light grey. The horizontal bar in the middle appears to progress from light grey to dark grey, but is in fact just one colour ("Optical illusion," 2013).

In brief, the theory of predictive coding postulates that forward, bottom-up connections propagate prediction error signals with (unexpected) sensory information about the stimulus from 'lower' (sensory) areas to areas 'higher' in the cortical hierarchy (Friston, 2010; Rao & Ballard, 1999). In this framework, the top-down backward connections pass predictions based on an internal (generative) model about the stimulus to lower-level sensory areas to minimize sensory prediction error (by selectively sampling the stimulus array) and to induce behavioural responses (Clark, 2012; Friston, 2010). Predictive coding and the Bayesian brain theory are incarnations of a more generalized theory known as the free energy principle of brain functioning (Friston, 2010). The large scale organisation of the human brain has been proposed as hierarchical in the cortex and organized along the rostro-caudal gradient, going from (primary) sensory-, to association-, to higher-level areas (Kiebel, Daunizeau, & Friston, 2008). The hierarchy mirrors the temporal structure of an agent's environment: the faster fluctuations in the environment activate the more sensory areas whereas increasingly slow fluctuation activate ever higher level areas (Kiebel et al., 2008).

This principle allows some neuroimaging evidence to be recast in a different light and makes specific predictions for future studies. For instance, if tasks are too constrained, the cognitive set will influence the activations. Staying with the example of the FFA functionality, it has recently been shown that when participants predict that they will be shown a face, the FFA area will respond just as much when participants are, against their expectations, shown a house as when they are shown, in line with their expectations, a face (Egner, Monti, & Summerfield, 2010). Moreover, research by Gallant et al. (Çukur, Nishimoto, Huth, & Gallant, 2013) shows that if people are instructed to attend to humans vs. cars in natural scenes, vast parts of the occipito-temporal and fronto-parietal cortex become active to the attended category. Thus, expectations about the experiment can greatly influence activations in neuroimaging experiments.

Another example are studies on the neural correlates of facial attractiveness that show that not only are reward related areas like the orbitofrontal cortex and striatal regions activated by viewing attractive faces, but also that FFA activity is increased when participants view more attractive faces (Iaria, Fox, Waite, Aharon, & Barton, 2008). This latter FFA activity is then sometimes interpreted as coding facial attractiveness by checking whether the 'face coding' area FFA activates particularly highly when viewing more attractive faces, which are more symmetrical and closer to averaged faces (Rhodes, 2006). However, in a hierarchical predictive coding framework, in the absence of context dependent effects, one would expect *reduced* activity in areas to attractive faces that diverge less from the average face and thus the internal model that people have of faces. Increased activity in the FFA activity when viewing attractive faces thus fits with the area's hypothesized role as a face expectation area and its activity might better be attributed to the surprising appearance of a face that people in their normal environment are statistically less likely to encounter – like an attractive face or unattractive face.

Finally, repetition suppression (sometimes called neural adaptation) fMRI paradigms have been argued to provide evidence for the predictive coding hypotheses. In repetition suppression, task stimuli are shown repeatedly and the cortical areas that show more reduced activity after repeated exposure are said to be more involved in the processing the stimulus. The exact neural mechanism of this effect are unknown (Grill-Spector, Henson, & Martin, 2006). Recently, some people have argued that this can be explained in terms of predictive coding. After being exposed to a stimulus it is more likely to reencounter the same stimulus than a different one- it is more likely and so the sensory prediction errors resulting from the repetition does not cause as much activity: it is already explained away by predictions (Clark, 2012). This view is supported by the finding that the repetition suppression effect is itself reduced if the repetition becomes unlikely, pointing to the role of perceptual expectations/predictions influencing activity (Summerfield, Trittschuh, Monti, Mesulam, & Egner, 2008), but these findings have been challenged (Kovács, Kaiser, Kaliukhovich, Vidnyánszky, & Vogels, 2013).

In the next section, I summarize the studies in the thesis.

1.7. Summary and the current thesis

The thesis' experimental chapters are ordered hierarchically, from high to low level behaviours, in the sense that it goes from the very high-level social behaviour required for interactive cooperation (trust, social exclusion) over spatial perspective taking needed during social interactions to agency detection required to initiate social behaviours.

1.7.1. Social Inclusion and Trust

In Chapter 2, I test the effects of social inclusion and exclusion on behaviour in a trust game. Social exclusion is very distressing and this distress might be an adaptive response to evolutionary pressures, mimicking physical pain. There are two possible responses when people are socially excluded: people might try to regain favour of the group i.e. the perpetrators of social exclusion, by putting forward even more trust and signalling the wish to cooperate (which might make them gullible to exploitation) or alternatively it might make them more cautious about being exploited by, and less trusting of, the people who have socially excluded them. In other words, the research question was whether participants will become more gullible or cautious of exploitation after social exclusion. I attempted to investigate this question by having participants included or excluded in virtual ball throwing game and then subsequently asking them to play an economic trust game against either the people who included or excluded them or strangers they had not met before. I investigated this by testing how much participants trust people who have socially included or excluded them. Inclusion and exclusion were manipulated using Cyberball (a virtual ball game) and, after playing Cyberball, participants subsequently played trust games. In a Reputation-Group participants played trust games with players from Cyberball; in the No-Reputation-Group, participants played with strangers.

1.7.2. The development of perspective taking

In chapter 3, I compare the neural correlates of taking the perspective of another person in adolescents and adults. Our everyday actions are often performed in the context of a social interaction. We previously showed that, in adults, selecting an action on the basis of either social or symbolic cues was associated with activations in the frontoparietal cognitive control network, while the presence and use of social vs. symbolic cues was in addition associated with activations in the temporal and medial prefrontal cortex (mPFC) social brain network. Here I investigated developmental changes in these two networks. Fourteen adults (age 21-30) and 14 adolescents (11-16) performed an adapted version of the Director Task described above. They followed instructions to move objects in a set of shelves. Interpretation of the instructions was conditional on the point of view of a visible “director” or the meaning of a symbolic cue, and the number of potential referent objects in the shelves. This study attempts, therefore, to show developmental differences in domain-general and domain-specific PFC activations associated with action selection in a social interaction context.

1.7.3. The effective connectivity of perspective taking

In chapter 4, I perform a connectivity analysis on the same data using Dynamic Causal Modelling (DCM) on the adult data from the previous chapter. Previous studies have shown that taking into account another person's perspective to guide decisions is more difficult when their perspective is not congruent with one's own compared to when it is congruent. I used DCM for fMRI to investigate effective connectivity between prefrontal and posterior brain regions that are activated while participants perform the Keysar task. Using a new procedure to score model evidence without computationally costly estimation, we conducted an exhaustive search for the best of all possible models. The results elucidate how the activity in the areas from our previously reported general linear model analysis of the adult data (Dumontheil et al., 2010) are causally linked and how the connections are modulated by both the social as well as executive task demands of the task. I interpret the results in terms of hierarchical predictive coding in chapter 6.

1.7.4. The effective connectivity of animacy perception

In Chapter 5, I used data from the Human Connectome project (HCP) to investigate the neural mechanisms of animacy perception. Biological agents are the most complex systems humans have to model and predict – a computationally demanding task. In predictive coding, high-level cortical areas inform sensory cortex about incoming sensory signals and unpredicted sensory information is passed forward to higher-level areas. Predictions about animate motion – relative to mechanical motion – should increase signal passing from lower level sensory area MT+/V5, responsive to all motion, to higher-order posterior superior temporal sulcus, selectively activated by animate (biological) motion. I tested this hypothesis by investigating effective connectivity in a large-scale fMRI dataset from the Human Connectome Project. Participants viewed animations of triangles that were either animate (moving intentionally), or inanimate (moving in a mechanical way). We hypothesized that the forward connectivity from V5 to pSTS increased and pSTS's inhibitory self-connection decreased, when viewing intentional motion versus inanimate motion. We speculate that animate motion prediction error may underlie enhanced attention and the phenomenological states of agency perception.

2. Experimentally induced social inclusion influences behaviour on trust games

Two interesting topics of study in social neuroscience are social exclusion and trust. Humans are profoundly gregarious animals who, in order to survive, must avoid exclusion when their groups try to increase cohesion (Williams, 2007, 2011). Trust is inextricably linked with cooperation in that it is a prerequisite for successful cooperation, and its establishment remains one of the big questions in behavioural science (Pennisi, 2005, but see West, El Mouden, & Gardner, 2011). Here we investigated how social exclusion influences trust.

This chapter is based on the following publication:

Hillebrandt, H., Sebastian, C. & Blakemore, S. (2011). Experimentally induced social inclusion influences behaviour on trust games. *Cognitive Neuroscience*, *2*(1), 27-33.

2.1. Introduction

Humans strive to be included within social groups and being excluded is profoundly distressing (Lieberman & Eisenberger, 2009). Previous studies have shown that social exclusion has wide ranging consequences including social pain (which shares neural substrates with physical pain), stress, sadness or anger; and that it reflexively threatens fundamental needs such as self-esteem, belonging and perceived control (Williams, 2007, 2011). Functional magnetic resonance imaging (fMRI) studies investigating the neural bases of social rejection have implicated several regions involved in socioaffective cognition, including the dorsal, ventral and subgenual anterior cingulate cortex (ACC), right ventrolateral prefrontal cortex (VLPFC), amygdala and insula (Eisenberger et al. 2003, 2007a; Somerville et al.2006; Kross et al. 2007; Masten et al. 2009).

Most notably, the finding that social pain is neurophysiologically similar to physical pain underscores why social exclusion can literally hurt. A common response to social exclusion is to behave in ways that improve the inclusionary status of the individual by, for example, acting in ways that increase acceptability to others and thus strengthening interpersonal bonds (Williams, Cheung, & Choi, 2000). Maner and colleagues found that after social exclusion, participants express greater interest in making new friends, show an increased desire to work with others, form more positive impressions of novel social targets, and assign greater rewards to new interaction partners (Maner, DeWall, Baumeister, & Schaller, 2007). Other studies have shown that after social exclusion, individuals are also more likely to donate money to a student organization (Carter-Sowell, Chen, & Williams, 2008), are more likely to cooperate (Williams & Sommer, 1997), and engage in more nonconscious mimicry towards subsequent interaction partners (Lakin, Chartrand, & Arkin, 2008).

Endocrinologically, one study found that after exclusion, levels of progesterone, a hormone that has been linked to social-affiliative motivation, depend on participants' levels of social anxiety and rejection sensitivity (Maner, Miller, Schmidt, & Eckel, 2010). However, there is also evidence that the behavioural consequences of socially excluded individuals are selective and differ depending on the interaction partners: one study showed that participants evaluated the person who had rejected them unfavourably and reduced the cash reward of that person (Maner, DeWall, Baumeister, & Schaller, 2007). An open question is thus whether socially excluded people become more gullible and socially susceptible because they want to improve their inclusionary status or whether they become more cautious to avoid exploitation in order to avoid costs in settings in which cooperation is necessary but potentially risky (Williams, 2007).

Here, we investigated the possibility that social exclusion makes people selectively less trusting of the specific person who excluded them in order to avoid falling prey to exploitation, and conversely makes them more trusting of the person who included them. Here, we created experimental situations of social inclusion and exclusion and measured their impact on reputation with game theoretical trust games. In the trust game, participants have to decide how much of an initial monetary endowment to invest in another player (Berg, 1995). The aim was to investigate whether exclusion leads participants to mistrust (as measured by money invested) individuals who have previously excluded them; and conversely whether a brief episode of social inclusion would increase trust in a subsequent economic game.

We employed Cyberball (Williams, Cheung, & Choi, 2000), an ostensibly 'online' ball-tossing game, to include or exclude participants who subsequently played iterated trust games. We assigned participants to two groups. In the Reputation group participants played the trust games with the fictional players with whom they had previously played the Cyberball games. In the No Reputation group participants played the trust games with fictional players not previously encountered (see Figure 1). We hypothesized that participants would show more trust towards fictional players who previously included them than to players who had excluded them in the ball game. We hypothesized that this effect would be based on social reputation and not merely based on mood induced by inclusion or exclusion, so that a difference in trust after the inclusion relative to exclusion manipulation would only be seen in the Reputation group.

2.2. Methods

2.2.1. Participants

We recruited 24 female participants in order to control for gender effects and so that everyone would 'play' with players of her own gender (for gender effects in trust games, see Johnson & Mislin, 2008). Participants were randomly divided into two groups and completed the two subtest-scale Wechsler Abbreviated Scale of Intelligence (WASI): the Reputation group (*N*=12; mean/SD age = 20.50/3.09; mean/*SD* IQ = 118.25/11.45), and the No Reputation group (*N*=12; mean/*SD* age = 21.42/2.91; mean/*SD* IQ = 119.42/11.64). The two groups showed no significant difference in age (*t*(22)= -0.75; *p* = 0.46) or IQ (*t*(22)= -0.25; *p* =0.81).

2.2.2. Design and procedure

The design was a 2x2 factorial with within-subjects factor Condition (Inclusion, Exclusion) and between-subjects factor Group (Reputation, No Reputation). Participants first played a round of Cyberball during which they were either included or excluded. This was immediately followed by an iterated trust game with three rounds (Figure 1).

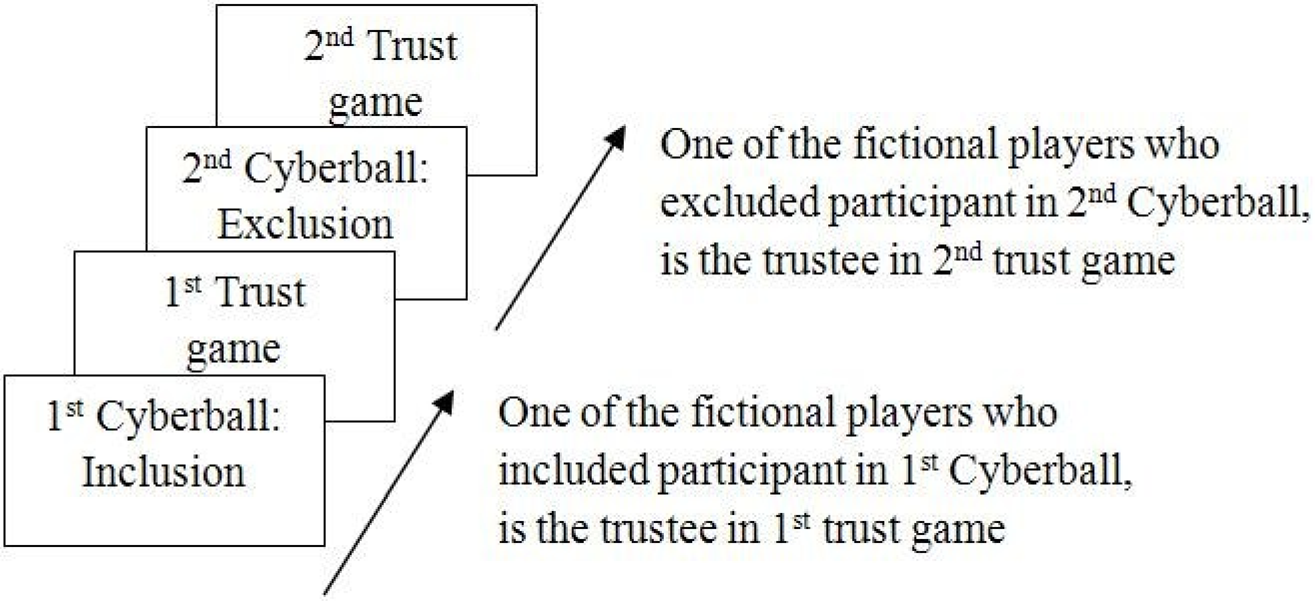

2nd Trust game
2nd Cyberball: Exclusion
1st Trust game
1st Cyberball: Inclusion
One of the fictional players who excluded participant in 2nd Cyberball, is the trustee in 2nd trust game
One of the fictional players who included participant in 1st Cyberball, is the trustee in 1st trust game

Figure 1: Example of the experimental setup for the Reputation group. Participants play inclusion or exclusion Cyberball first and then re-encounter one of the fictional players from Cyberball in the subsequent trust games. Each of the two iterated trust game has three rounds. In the No Reputation group, participants play against different fictional players in every game.

The overall sequence was repeated, so that all participants were included and excluded in different Cyberball games once. After each Cyberball game, participants played a trust game with three rounds; sequence order was counterbalanced between participants. Participants were first given detailed instructions about the two games they would play.

Participants were told that they would play two online games with other players whose photographs they would see throughout the game, together with an identifying symbol (in reality participants played against algorithms). Photographs were taken from those included with the Cyberball software (Williams, Cheung, & Choi, 2000) and supplemented with photographs of female researchers working at our institute (who would not be present). To increase credibility, we took a photograph of each participant at the start of the experiment and informed them that this photograph and a symbol for recognition would be displayed throughout the game to the other players. In order to minimize effects of social desirability, participants were told that they would not see the other players in person during or after the experiment. After reading through the instructions for the trust game, and before the games commenced, in order to create a realistic atmosphere, participants were told to wait, since the 'other players were not ready to play yet'. Participants were left alone for 2 minutes and then notified via chat that all players were ready, at which point they were shown a screen with Cyberball instructions.

2.2.3. Cyberball

Cyberball (Williams, Cheung, & Choi, 2000) was introduced as a way of testing "the effects of practicing mental visualization on task performance". We used this game to induce social exclusion and inclusion. Participants played Cyberball twice (one inclusion and one exclusion game, with game order counterbalanced between participants). Each game was played with two novel players. Participants saw the photographs of the other players and a symbol to aid recognition (see Figure 2).

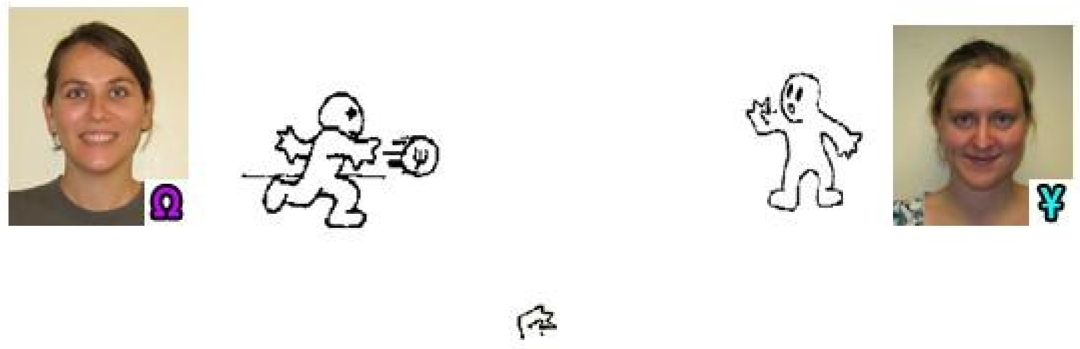


Figure 2. Screenshot of a Cyberball game interface. Participants are represented by a cartoon hand at the bottom of the screen, and photographs of the other two (fictional) players are shown on either side. The ball is thrown 30 times in total. In the inclusion condition, participants receive the ball approximately one third of the time. In the exclusion condition, after two initial throws from the other players, participants were not thrown the ball again.

In both inclusion and exclusion there were 30 throws of the ball in total. In the exclusion game, after two initial throws, participants were not thrown the ball again. In the inclusion game participants were thrown the ball approximately half of the time by each of the other two players. The photographs used in the Cyberball games and trust games were also counterbalanced between participants, as well as the side on which the trust game player's photograph appeared in the Cyberball game (left, right).

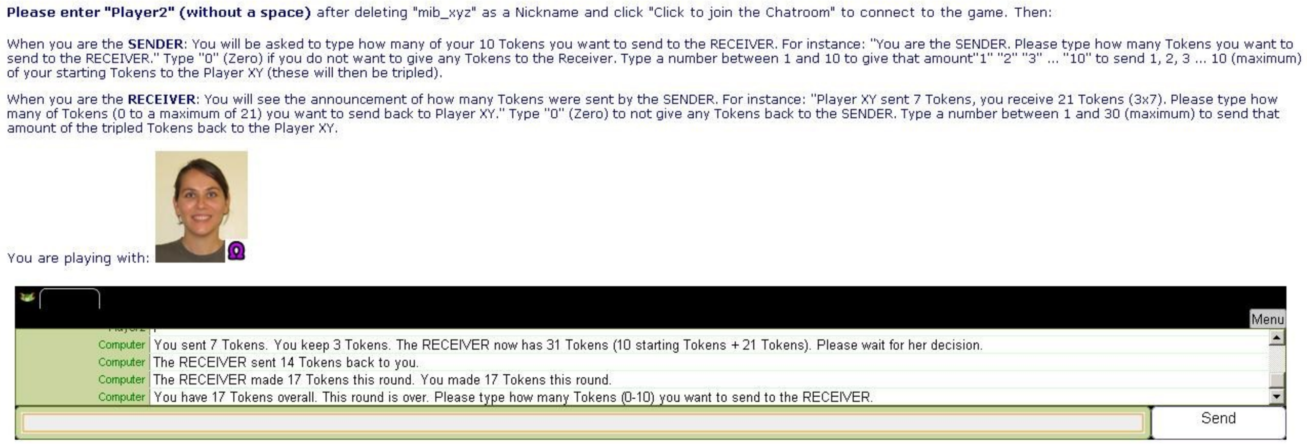
2.2.4. Trust game

Immediately following each Cyberball game, participants played a 3-round trust game with one of the players with whom they had just played the Cyberball game (in the Reputation group) or with other players they had not encountered before (in the No Reputation group). The game's description was neutral in order not to bias participants' reactions; the word 'trust' was omitted. Participants were told that they would play a game with another player and they would both have the chance to win tokens. Participants were told that tokens would be converted into real money at the end of the game at a conversion rate that would be revealed after the game. Participants read that there would be "a number of" rounds, so that the game would be interpreted as an iterated game with an indefinite number of rounds.

Participants were informed that in each game one player is the ‘sender’ (the trustor in game theory terminology) and one the ‘receiver’ (the trustee), and that it is decided at random who the sender and the receiver is at the beginning of each game. In reality, the participant was always the sender (trustor) in both trust games.

In the trust game, each player starts each round with 10 tokens. The sender can decide to keep all her starting tokens or send any number of her tokens to the receiver. The number of tokens the sender sends to the receiver is tripled. The receiver receives the tripled number of tokens and can then decide either to keep all the tokens for herself (to defect in game theory terminology) or send any number of her accumulated tokens back to the sender (to cooperate). The number sent back is not tripled.

Participants were informed that they would play the game in a virtual chat room (see Figure 3), and that the computer would record their decisions in the form of their chat input and mediate the game with the other player in another chat room. In order to decrease effects of social desirability, participants learned that the experimenter could not see the individual decisions they made, but only the final result of how many tokens each participant won. Participants were given examples of a round.



Please enter "Player2" (without a space) after deleting "mib_xyz" as a Nickname and click "Click to join the Chatroom" to connect to the game. Then:

When you are the **SENDER**: You will be asked to type how many of your 10 Tokens you want to send to the RECEIVER. For instance: "You are the SENDER. Please type how many Tokens you want to send to the RECEIVER." Type "0" (Zero) if you do not want to give any Tokens to the Receiver. Type a number between 1 and 10 to give that amount"1" "2" "3" ... "10" to send 1, 2, 3 ... 10 (maximum) of your starting Tokens to the Player XY (these will then be tripled).

When you are the **RECEIVER**: You will see the announcement of how many Tokens were sent by the SENDER. For instance: "Player XY sent 7 Tokens, you receive 21 Tokens (3x7). Please type how many of Tokens (0 to a maximum of 21) you want to send back to Player XY." Type "0" (Zero) to not give any Tokens back to the SENDER. Type a number between 1 and 30 (maximum) to send that amount of the tripled Tokens back to the Player XY.

You are playing with:

Computer You sent 7 Tokens. You keep 3 Tokens. The RECEIVER now has 31 Tokens (10 starting Tokens + 21 Tokens). Please wait for her decision.
Computer The RECEIVER sent 14 Tokens back to you.
Computer The RECEIVER made 17 Tokens this round. You made 17 Tokens this round.
Computer You have 17 Tokens overall. This round is over. Please type how many Tokens (0-10) you want to send to the RECEIVER.

Figure 3. Screenshot of the trust game chat interface. In the Reputation group participants saw a photograph of one player with whom they just played Cyberball, whereas in the No Reputation group they saw someone they had not encountered yet. Participants were ‘allotted’ the sender role and asked to make a decision about how many of the 10 tokens they wanted to send to the other player. After their decision, they received a confirmation statistic telling them how many tokens they sent, how many tokens they keep, how many tokens the receiver now has, and a request to wait for the receiver’s decision. After a short delay, they received a message and statistics about how many tokens the receiver sent back to them, and how many tokens the receiver and the participant made in that particular round and in the whole game.

In the chat room, participants in the Reputation group saw a photograph of one player with whom they had just played Cyberball, whereas participants in the No Reputation group saw someone they had not previously played with. Participants read that they were allotted the sender role and asked to make a decision about how many of the 10 tokens they wanted to send to the other player (see Figure 3 for details). The number of tokens sent back to the participant was randomly jittered, so that both players ended up with the same amount of tokens or the participant ended up with two tokens more than the other player. This jitter was included to increase the realism of the game by avoiding always having the same outcome of a fair split.

After the games, participants filled in a questionnaire and were debriefed; we emphasized that they had not played with real people and thus were not really excluded. Finally, participants were reimbursed a set amount (£7.50). The experiment took approximately one hour.

2.2.5. Manipulation check

Participants completed a questionnaire containing statements about how excluded and ignored they felt in the different games, and to what extent they agreed with these items on a Likert scale from 1 ('not at all') to 5 ('very much so'). Participants were also asked what percentage of throws was directed to them in the two Cyberball games, and to rate the "inclusion of other in the self scale" (Aron, Aron, & Smollan, 1992) to find out how close they felt to the other players.

2.3. Results

All participants noticed that they did not receive the ball in the exclusion game as much as in the inclusion game ($t(23) = 8.29$, $p < 0.001$) and felt significantly more ignored ($t(23) = 10.94$, $p < 0.001$) and excluded ($t(23) = 9.80$, $p < 0.001$) in the exclusion game. They felt significantly closer to the players in the inclusion game than to the players in the exclusion game, as assessed by the inclusion of other in the self scale ($t(23) = 2.25$, $p = 0.04$).

We measured trust by calculating how many tokens the participant sent to the other player overall during the three rounds, for each of the groups. We carried out mixed model repeated measures 2x2 ANOVA with between subjects factor Group (Reputation vs. No Reputation) and within subjects factor Condition (Inclusion vs. Exclusion). There was no main effect of Condition (Wilks' Lambda = 0.99, F (1,22) = 0.26, MSE = 2.08, p = 0.62; partial η2 = 0.01), or Group (F (1,22)= 0.46, MSE = 52.08, p = 0.46; partial η2 = 0.03), but there was a significant interaction between Group and Condition (Wilks' Lambda = 0.80, F (1,22) = 5.52, MSE = 44.08, p = 0.03, partial η2 = 0.20; see Figure 4).

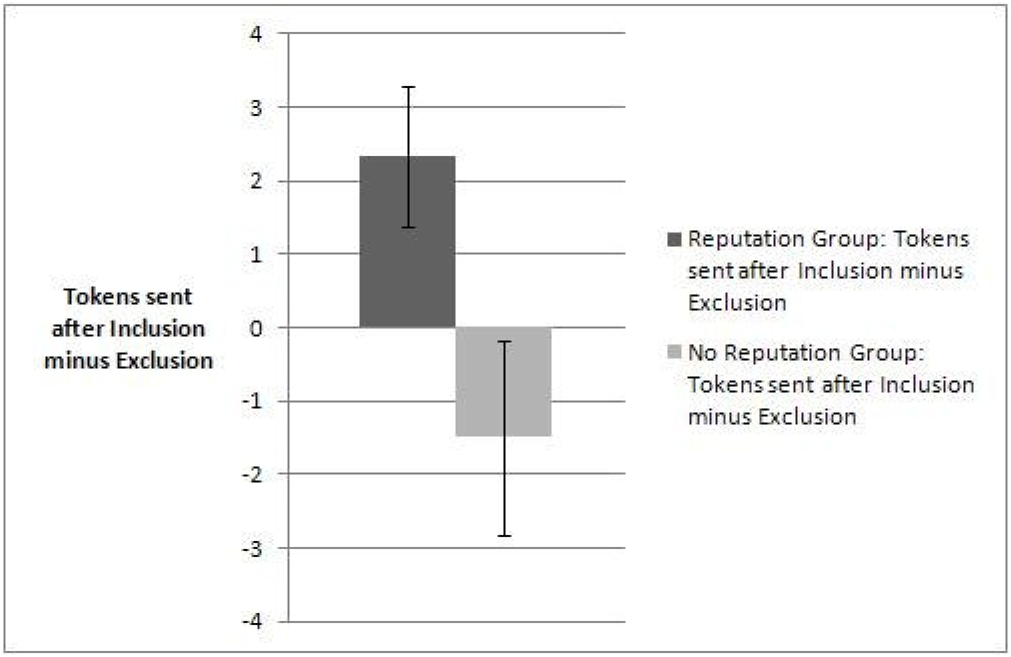


Figure 4. Graph showing the significant interaction of Group by Condition: In the Reputation group more tokens were sent after inclusion than exclusion, whereas in the No Reputation group there was no significant difference. Error bars show standard error of the mean.

A paired-samples *t*-test revealed that the Reputation group sent significantly more tokens in the trust game after inclusion (*M* = 17.75, *SD* = 6.37) than after exclusion (*M* = 15.42, *SD* = 5.42) in the Cyberball game (*t* (11) = 2.46, *p* = 0.03, *d* = 0.71). However, this was not the case for the No Reputation group who showed no significant difference between the number of tokens sent after inclusion (*M* = 13.75, *SD* = 7.83) and exclusion (*M* = 15.25, *SD* = 8.40; *t* (11) = -1.13, *p* = 0.28, *d* = -0.32).

2.4. Discussion

This study showed that participants trusted players who previously included them more than they trusted players who excluded them in an online ball tossing game. Since the difference in trust after the inclusion and exclusion manipulation was only seen in the Reputation group, this effect must have been social in the sense that different levels of trust were specific to fictional players who included or excluded the participant. Therefore, the present findings indicate a true effect of reputation, over and above any general (mood) effects caused by being included or excluded prior to the trust game. Our data do not support the proposal that social exclusion renders victims of exclusion more gullible and easy to manipulate in general, but rather that participants are as cautious towards those who excluded them as to strangers. Moreover, the data suggest that social inclusion increases trust. While there was less trust after exclusion than after inclusion in the Reputation group, the level of trust after exclusion was not different from the level of trust that strangers were met with (in the No Reputation group) - a level very close to 50.88% (corresponding to 15.26 tokens in our study), which was reported in a recent meta-analysis of 84 trust games (Johnson & Mislin, 2008).

Our findings are in line with research on the influence of reputation on trust. For instance, one recent trust game study found that trustors were more likely to trust when they played with a trustee whose history showed that he was trustworthy in the previous trust game he had played with other people (Bracht & Feltovich, 2009). In another study, participants cooperated more with, and donated more money to, the other player in an Iterated Prisoner’s Dilemma game when playing with a friend rather than with a stranger (Majolo et al., 2006). Our data suggest that a short and relatively superficial social encounter is sufficient to increase trust. The functional relationship between social inclusion/exclusion and trust thus seems to be that after exclusion it is better for people to be cautious about being exploited by those who excluded them. Social inclusion, on the other hand, might signal willingness for cooperation, which creates a (minimal) in-group whose members should reciprocate and should be trusted in order to solidify established relationships. This is in line with research showing that trusting is more likely in situations of low uncertainty, and that individuals of small, rather than large, social distance are more likely to be trusted (Goto, 1996).

Moreover, because humans strive to be included within groups, social inclusion is commonly described as feeling good and social affiliation is rewarding. One possibility is that our findings are due to increases in the neuropeptide oxytocin: it has been suggested oxytocin is released after positive social interactions, such as social support or social proximity (Heinrichs, von Dawans, & Domes, 2009). Interestingly, one study showed that intranasal administration of oxytocin increases trust during a trust game in humans (Kosfeld, Heinrichs, Zak, Fischbacher, & Fehr, 2005). Another fMRI study found that during trusting in a trust game participants showed activation in the septal area (and the adjoining hypothalamus), which is responsible for release of oxytocin and the related neuropeptide vasopressin, and is thus linked to social attachment behaviour (Krueger et al., 2007). In our study, re-encountering the people who socially included the participants might have increased levels of oxytocin and thus increased trust. Trusting others during iterative trust games is associated with activity in the striatum and has been proposed to be similar to reinforcement learning (King-Casas et al., 2005). Interestingly, another study (Delgado, Frank, & Phelps, 2005) built upon these findings and showed that different prior social information (positive, neutral and negative) about trustees modulates the activity within the brain's reward circuitry. The caudate nucleus differentiated between positive and negative feedback (i.e. trust repayment) only for neutral (and weakly for negative) described trustees, but not for the positively described trustees, suggesting that prior social information (like inclusion) can reduce reliance on feedback mechanisms in the neural networks associated with trial-and-error reward learning (Delgado, Frank, & Phelps, 2005). Behaviourally this was mirrored by the finding that, despite equivalent reinforcement rates (i.e. repayment of trust), participants were more likely to trust the trustee about whom they had read positive information. This is similar to our findings, which show that participants place more trust with individuals who have previously included them in an online game even though the repayment of the trust was similar after inclusion and exclusion.

There has recently been a call for creating and examining psychologically meaningful social situations in cognitive neuroscience (Todorov, Harris, & Fiske, 2006). The current study was designed such that the independent variable was psychologically meaningful and externally valid, and the dependent measure was also embedded in that same continuing psychologically meaningful situation. The social context was manipulated experimentally with different social situations and its impact on behaviour was measured with game theoretical dependent measures embedded in this social situation. In sum, our results suggest that financial decisions such as whether to trust someone with one’s money can be influenced by a previous brief online social encounter with that person.

In order to make decisions about whether one can trust another person, one has to take the perspective of another person. In the next two chapters, we investigate the neural mechanisms of perspective taking.

3. Developmental Differences in the Control of Action Selection by Social Information

This chapter is partly based on a manuscript in which I share first authorship:

Hillebrandt, H.*, Dumontheil, I.*, Apperly, I. A., & Blakemore, S.-J. (2012). Developmental Differences in the Control of Action Selection by Social Information. *Journal of Cognitive Neuroscience*, 24(10), 2080–2095. doi:10.1162/jocn_a_00268 * indicates shared first authorship

3.1. Introduction

How is an appropriate action selected among many possibilities? How do we decide to pick a particular fruit out of many on offer at the supermarket? The prefrontal cortex (PFC) is thought to support the integration of information from the current environment (the smell, colour, appearance and identity of melons on stand), and internal information generated more or less remotely in time (the plan to invite friends for dinner, the decision to have melon as a starter, but also the value associated with melons compared with other fruits), in order to orient attention appropriately (towards the stand and the most appetizing looking melon) and select an action appropriate with the current goals (picking up the selected melon) (e.g. Baron-Cohen, Leslie, & Frith, 1986; Burgess, Gilbert, & Dumontheil, 2007; Fuster, 2000; Koechlin & Summerfield, 2007). Using visual search paradigms, previous research has investigated the selection of targets combining different properties in arrays of simple stimuli (Booth et al., 2003; Davis & Palmer, 2004; Humphreys, Allen, & Mavritsaki, 2009). Results suggest a dissociation between bottom-up (spontaneous orientation towards a stimulus) and top-down (intentionally driven by knowledge, expectations and goals) processes of visuospatial selective attention (Beck & Kastner, 2009; Hahn, Ross, & Stein, 2006). However, few studies have attempted to use more complex stimuli.

A large number of our everyday actions are performed in a social interactive context. If you are paying at a shop checkout, you need to perform a series of actions that ensures the shopkeeper knows whether you wish to pay by cash or credit card, and so on. These interactions are sometimes based on verbal communication, and at other points on communicative gestures. For example, making sure the shopkeeper can see you have taken your credit card out of your wallet informs the shopkeeper that you want to pay by card. This type of interaction relies on an understanding of other people's mental states, also called theory of mind or mentalising (Frith & Frith, 2007). Thus, we often need to use theory of mind to make top down decisions and select actions that are appropriate to the inferred mental state of the people we interact with in complex real-world situations. The combination of domain-general processes supporting selective attention, action selection and cognitive control, and cognitive processes that may be specific to social information and mentalising, is the focus of the present study. The PFC is involved in both top down action selection (Burgess et al., 2007) and mentalising (Frith & Frith, 2007), in particular in ill-structured or novel situations (Apperly, 2011). Following previous work in adults (Dumontheil, Küster, Apperly, & Blakemore, 2010) the aim of the current study was to investigate the development of these domain-general and domain-specific processes during adolescence, using functional magnetic resonance imaging (fMRI) and a paradigm that permits the comparison of action selection in a social vs. non-social context.

In the current study, we adapted a computerised version of the Director task (Apperly et al., 2010; Dumontheil, Apperly, & Blakemore, 2010) for fMRI (Dumontheil, Küster, et al., 2010).

While the previous studies using this paradigm were designed to probe the use of social cues (perspective taking) in a natural way via assessing differential error rates (which are naturally quite high), the present version included extensive task instructions and practice in order to minimise error rates, and compare fMRI data across groups performing at consistently high levels of accuracy. Our goal was to use this task variant to assess the development between adolescence and adulthood of the neural substrates associated with: (1) the selection of an appropriate action in the context of alternative options proposed in a complex stimulus; (2) the presence of social information vs. symbolic cues as part of the stimuli; (3) the use of social cues as opposed to symbolic cues for the selection of the appropriate action from the alternative options. Note that, while beyond the scope of the current study, the computation of the value of different objects or actions may be another important aspect of action selection, and may be affected by personal and social factors developing during adolescence.

If mentalising is a specialised and domain-specific cognitive process (e.g. Leslie, 2005; Sperber & Wilson, 2002; Stone, Baron-Cohen, & Knight, 1998; see Apperly, 2011 for discussion), then we might expect differential performance and the recruitment of regions of the social brain network (Brothers, 1990; Frith & Frith, 2007) when the guiding information is of social nature compared to more arbitrary symbolic stimuli. This social stimuli condition might thus recruit both domain-general action selection resources and domain-specific resources that are involved in processing social information and mentalising. Such a pattern has previously been observed in adults, with non-overlapping brain regions implicated in response selection and belief attribution during a belief attribution task (Rebecca Saxe, Schulz, & Jiang, 2006; see also Saxe, Carey, & Kanwisher, 2004).

Participants followed auditory instructions to move objects in a set of shelves. A 2 × 2 factorial design with the factors Director (Director Present vs. Director Absent) and Object (3-object vs. 1-object) was employed. In the Director Present condition, two directors were shown on the display, one female and one male. One of them stood behind the shelves, facing the participant, while the other stood on the same side of the shelves as the participant. In the 3-object condition, participants needed to use the social cues, i.e. the position of the speaking director, to select and move the appropriate objects (see **Figure 1A**). The instructions in these blocks referred to an object that was one of three exemplars in the shelves (the ball). Two of these objects could correspond to the heard instruction ("Move the large ball up") depending on the director's viewpoint. The largest ball (or equivalent) was always located in a closed shelf (not visible from the back), while the second largest ball was located in an open shelf. Importantly, when considering the identity of the director (male voice), only one of the objects was the appropriate response (the football); the other object corresponded to the other viewpoint and was thus a distractor (the basketball**)**. The third object was another type of distractor that did not fit any of the perspectives (the tennis ball). On half of the Director Present 3-object trials the perspective of the director issuing the instruction was different from that of the participant; on the other half the director's and participant's perspectives were the same. This varied on a trial-by-trial basis, thus participants needed to take into account the director's perspective on every trial. In Director Present 1-object trials, there was no need to take into account the director's perspective to identify the correct object (e.g. "Move the turtle left"), as there were no distractors or other referents, and so this resembled a bottom-up, visual pop-out as opposed to a visual search (Buschman & Miller, 2007). The Director Absent condition was logically equivalent, but the directors were replaced by symbolic cues (see **Figure 1B**).

We previously obtained results on this task in a group of adult participants (Dumontheil, Küster, et al., 2010). Selection of an appropriate action when faced with alternatives (3-object vs. 1-object contrast, collapsed across Director Present and Director Absent conditions) was associated with domain-general bilateral brain activations located primarily in the fronto-parietal cortex, with additional activations in the inferior temporal cortex. Processing of social (Director Present) vs. symbolic information (Director Absent) was associated with specific activations in the superior dorsal MPFC and superior temporal sulcus. Finally, using perspective taking in this communicative context (Director × Object interaction), which required participants to think not only about what the other person sees but also about his/her intentions, led to further recruitment of superior dorsal MPFC and the left middle temporal gyri, extending into the temporal pole. The focus of the current study was the difference between the adolescent and adult groups in these three contrasts of interests: (1) 3-object > 1-object; (2) Director Present > Director Absent; and (3) Director × Object. In line with the adult data, we expected activation in the cognitive control frontoparietal regions for the first comparison, and activation in the social brain network in the latter two comparisons, which contrast the presence of social vs. symbolic information, and the specific perspective taking requirement. In terms of developmental effects, for the first comparison, as the direction of changes in parietal and frontal cortex activations with age in cognitive control tasks is inconsistent in the literature and may be task dependent, we tested for both developmental increases and decreases in the 3-object vs. 1-object comparison. Regarding the second and third comparisons, the social brain has shown more reliable decreases between adolescence and adulthood in MPFC activations (Blakemore, 2008) and increases in temporal cortex activations (Blakemore, 2008; Burnett et al., 2011) in a variety of social cognition tasks. Our predictions thus aligned with this previous literature: we predicted decreased MPFC activation and increased temporal cortex activation with age in the Director Present vs. Director Absent comparison, and/or more specifically in the Director Present 3-object condition, which requires online use of perspective taking information. Regarding the MPFC, our prediction was focused on the dorsal MPFC on the basis of our previous results in adults, as well as the association between this region and mentalising judgements performed towards others (in this case the Directors) rather than the self.

3.2. Material and methods

3.2.1. Participants

Fourteen adult (mean age 24.9 years (standard deviation (*SD*) 3.0), age range 21.3–30.6) and 14 adolescent (mean age 14.0 (standard deviation (SD) 1.6), age range 11.6–16.8) right-handed female volunteers were included in the analyses (two additional adolescent participants performed poorly on the task due to malfunctioning headphones and were excluded). All participants spoke English fluently and had no history of psychiatric or neurological disorder. Adult participants or the parents of the adolescent participants gave informed consent and the study was approved by the University College London ethics committee. General ability was assessed using the two subtests format (Vocabulary and Matrix Reasoning) of the WASI (Wechsler, 1999). Estimated IQ normalised for age did not significantly differ between the adolescents (122 (± 11.7), 106-140) and adults (110 (± 9.8), 100-134) ($t(26) = .735$, $p = .47$).

3.2.2. Design and stimulus material

Stimuli consisted of sets of 4 × 4 shelves with objects located in half of the shelves. Five of the shelves had a grey background **(Figure 1)**. On each trial, participants were given instructions via headphones, by either a male or a female voice, to move one of the eight objects in the shelves to a different slot, either up, down, left or right (note that this was the participant's left or right). A 2 × 2 factorial within-subject design was used with the factors Director (Present vs. Absent) and Object (1-object vs. 3-object) varying between blocks.

3.2.3. Director factor

In the *Director Present* (DP) condition the display included two directors, one female and one male. In the *Director Absent* (DA) condition, there were no directors in the display (**Figure 1**). Instead, the letter "F" for female and "M" for male were shown beside the shelves. Below each of the letters there was either one transparent box, which indicated to participants that only objects in open shelves should be moved, or two boxes, one grey and one transparent, which indicated that there was no restriction on the participant's choice and all objects (both in open shelves and occluded shelves) could be moved. For example, in **Figure 1B**, if participants heard the male voice say: "Move the large ball up," they would need to reason that, since the M is above one clear box, they could only pick objects in clear shelves, and thus should ignore the basketball in the grey slot and move the football. These rules had precisely the same consequences as the position of the director in the DP blocks. In DP blocks the physical position of the director issuing the instruction varied on a trial-by-trial basis; similarly, in DA blocks the M/F rules changed on a trial-by-trial basis.

3.2.4. Object factor

Instructions in *1-object* blocks (e.g., in **Figure 1**, "Move the turtle left") referred to a unique target object (there was only one turtle), which was in an open shelf. Instructions in *3-object* blocks (e.g., "Move the large ball up") could refer to an object in a closed shelf (with a grey background) or an object in an open shelf, which could both be described with the same instruction (e.g. "large ball"). Which of the possible referents was in fact correct was determined by whether the director giving the instruction (identified as male or female by his/her voice) was at the back or front of the shelves (in DP), or whether the cues indicated that only objects in open shelves could be moved (in DA). This manipulation ensured that in DP 3-object blocks participants had to consider the director's perspective (which was different from their own perspective on 50% of the trials) in order to know which was the correct object to move. In DP 1-object blocks the director's perspective made no difference to the correct interpretation of his or her instructions, and thus participants could use their own perspective to select the appropriate object on all trials. In the DA condition, perspective taking was not involved.

There were 48 object-shelf configurations, each containing eight objects. Sets of three exemplars of the same object were used for 3-object trials (e.g. three drums). These objects differed in either size (large/small) or position (top/bottom) and were distributed so that the smallest/largest or topmost/bottommost object identified in the instruction was in a closed shelf and the second smallest/largest or topmost/bottommost object and the remaining object were in open shelves. Five additional unique objects were distributed in three grey-backed closed shelves and two open shelves. Those objects in the open shelves could be used for 1-object trials.

To move objects, participants used a trackball mouse, rolling the trackball with their thumb and pressing the left mouse button with the index finger of their right hand. On each trial, participants first moved the mouse cursor from the middle of the screen to the selected object, then clicked on the object and dragged it to the appropriate slot, before releasing the mouse button. Response times were calculated as the delay between the presentation of the visual stimulus and the pressing of the mouse button. Accuracy was measured on the basis of which object was moved.

On each trial the visual stimulus and the auditory instruction were presented over a period of 2.2 s, after which the display remained on the screen for another 3.8 s. Between trials a blank screen was shown for 200 ms. The task was programmed with Cogent 2000 and Cogent Graphics (www.vislab.ucl.ac.uk/cogent.php) implemented in Matlab 6.5 (Mathworks Inc., Sherborn, MA). Standardised instructions were read to participants, and included example stimuli in which they had to state which objects should be moved for the different directors and voices. A practice session including one block of each of the four conditions was run outside the scanner to ensure that participants understood the task and could perform it correctly. If a participant did not take into account the director's perspective appropriately, this was highlighted and the task requirements explained again. Participants also practiced using the trackball mouse ahead of scanning (see Dumontheil, Kuster, et al., 2010).

Participants performed three scanning sessions (two adolescent participants performed two sessions only because of time constraints). Each session consisted of 16 task blocks with four trials in each block. There were four types of experimental block: DA 1-object; DA 3-object; DP 1-object; DA 3-object. Each of the 48 object-shelf configurations was shown once in each block type, thus there were 12 blocks of 4 trials for each block type. Task blocks lasted 24.8 s and were preceded by an instruction screen presented for 2 s, which indicated whether the upcoming block was a DP or DA block. The order of the four block types was counterbalanced within and between sessions. A fixation baseline block lasting 20 s was included after each set of four task blocks.

3.2.5. fMRI data acquisition

3D T_1-weighted fast-field echo structural images and multi-slice T_2-weighted echo-planar volumes with blood-oxygen level dependent (BOLD) contrast (TR = 3 s; TE = 50 ms;

TA = 2.9143 s) were obtained using a 1.5 T MRI scanner (Siemens TIM Avanto, Erlangen, Germany). Functional imaging data were acquired in two or three scanning sessions lasting approximately 8 min 40 s each in which 174 volumes were obtained. The first 2 volumes of each session were discarded to allow for T_1 equilibrium effects. Each functional brain volume was composed of 35 axial slices with an in-plane resolution of 3 × 3 × 3 mm, positioned to cover the whole brain. A T_1-weighted anatomical image lasting 5 min 30 s was acquired after the first two functional sessions for each participant.

3.3. Data analysis

3.3.1. Behavioural data

Response times (RTs) and accuracy in all four conditions were recorded and analysed using a 2 (Age group: Adolescents vs. Adults) × 2 (Director: DA vs. DP) × 2 (Object: 1-object vs. 3-object) mixed model repeated measures ANOVA to investigate the effects of each task factor, the interaction between task factors and their interaction with age.

3.3.2. fMRI data

fMRI image preprocessing and analysis were carried out using SPM8 (Wellcome Department of Imaging Neuroscience, London, UK), implemented in MATLAB 7.8 (Mathworks Inc., Sherborn, MA). To correct for movement effects images were realigned with a 4th degree-B-spline interpolation. These realigned images were corrected for differences in acquisition times and were then normalised to a standard EPI template based on the Montreal Neurological Institute (MNI) reference brain. The resulting 3 × 3 × 3 mm images were finally spatially smoothed with an 8mm FWHM Gaussian kernel. Analyses of the movement parameters showed that translation within each session was < 3 mm in all participants. Mean movement per session were calculated for translations and rotations in each direction for each subject. Independent *t*-tests were used to compare the means of these values over the sessions between the two age groups. There was no significant difference for translations (all $t(26) < .43$, $p > .67$), nor for rotations (all $t(26) < .28$, $p > .78$).

For each participant the scanning sessions were treated as separate time series; statistical parametric maps were created and estimated using a general linear model for each time series (Friston et al., 1995). Included in the model were six boxcar regressors, modelling the instruction, fixation and four types of task blocks, plus one event-related regressor representing error trials. All regressors were convolved with a canonical haemodynamic response function and, together with regressors representing residual movement-related artefacts and the mean over scans, comprised the full model for each session. The data and model were high-pass filtered to a cut-off of 1/128 Hz. Parameter estimates calculated from the least mean squares fit of the model to the data were used in four pair-wise contrasts comparing each block type with the fixation baseline. These contrasts were entered into a condition × age group × participant flexible factorial design second-level analysis. The factor participant was included to model the repeated aspect of the data. Main effects of Object (3-object > 1-object) and Director (DP>DA and DA>DP), and the interaction between the two factors and with age group were determined using the *t*-statistic on a voxel-by-voxel basis. Statistical contrasts were used to create SPM{Z} maps thresholded at $p < .001$ at the voxel level and at family-wise error (FWE) corrected $p < .05$ at the cluster level (corresponding to a minimum cluster size of 82 voxels determined with SPM8). Activations that survived whole brain FWE correction at $p < .05$ are indicated. Analyses performed with age as a continuous regressor did not highlight any regions not observed with the age group analyses, and are thus not reported. All coordinates are given in MNI space. ROI analyses based on the main effect of DP > DA were further performed to test for orthogonal interaction effects between the Director, Object and Age group factors in the brain regions identified as responding to the social stimuli. Additional analyses explored possible continuous effects of age, and differences between the 7 youngest and 7 oldest adolescent participants and the adults. Note that these analyses are limited by the age range gap between 17 and 21 and underpowered by the small *N* of the adolescent sub-groups. Statistical threshold for the ROI analyses performed in SPSS was $p < .05$ (two-tailed).

3.4. Behavioural results

Accuracy was calculated for each of the four conditions in each session. Four sessions (from three subjects) were discarded from further analyses because of poor performance in one of the conditions (accuracy < 50%). Included in the analyses were thus 14 adults (13 with three sessions, 1 with two), and 14 adolescents (10 with three sessions, 3 with two, 1 with one). Accuracy and median RTs in correct trials were analysed using a 2 (Age group: Adolescent, Adult) ×2 (Director: DP vs. DA) × 2 (Object: 1-object vs. 3-object) mixed model repeated measures ANOVA.

Accuracy was higher in the DP than in the DA condition (main effect of Director, $F(1, 26) = 6.29$, $p = .019$), and higher in 1-object than 3-object blocks (main effect of Object, $F(1, 26) = 38.29$, $p < .001$)(**Figure 2A**). There was no main effect of age group ($p > .6$) and no interaction between age group and the task factors (all interaction *p*-values > .22). A similar pattern of performance was observed in terms of RTs. Participants were faster in the DP than DA condition ($F(1, 26) = 48.36$, $p < .001$), and in 1-object than 3-object blocks ($F(1, 26) = 465.16$, $p < .001$), and there was no main effect of age and no interaction with age group (all *p*-values > .24).

However, there was a significant interaction between Director and Object ($F(1, 26) = 15.69$, $p < .001$), reflecting a greater effect of Object in the DA condition (**Figure 2B**). Thus effects of Director and Object were observed on both measures of performance, but the adolescents' and adults' performance did not differ.

3.5. fMRI results

The four first level contrasts comparing each block type (Director (2) × Object (2)) to fixation were entered in a flexible factorial second level analysis including age group and participant as factors.

3.5.1. Object factor

A broad bilateral network of frontoparietal, occipital and inferior temporal regions showed increased BOLD signal in 3-object compared to 1-object blocks (**Table 1, Figure 3A**), i.e. when the participants had to identify a specific object to move among three exemplars of the same object type (e.g. one of several balls on **Figure 1**) as opposed to when there was only one exemplar of the object (e.g. the turtle on **Figure 1**). Greatest increases in BOLD signal were observed bilaterally in superior and inferior parietal lobules, superior frontal sulci and precuneus. Other regions included the medial superior frontal gyrus and anterior parts of the PFC in the middle frontal gyri bilaterally.

Object and Age group factors significantly interacted in frontoparietal regions in the left hemisphere. Adults showed increased BOLD signal in the 3-object vs. 1-object blocks, compared to adolescents, in the intraparietal sulcus and in a cluster extending from the precentral gyrus to the inferior frontal gyrus and insula (**Table 1, Figure 3B**). This significant interaction between Object and Age group reflected more bilateral activations in frontal and parietal regions in adults than adolescents (**Figure 3C**). Note that these two clusters remained significant when mean RT and accuracy for each condition and participant were entered as covariates in the 2nd level analyses (left frontal cluster: $Z = 4.84$, $p(\text{FWE}) < 0.001$, 326 voxels; left parietal cluster: $Z = 4.27$, $p(\text{FWE}) = 0.029$, 96 voxels).

Both younger and older adolescent sub-groups showed weaker 3-object than 1-object fronto-parietal activation than adults (*p*s < .05) and did not differ from each other (parietal region: $p > .5$; frontal region: $p = .066$, trend for greater activation in the younger adolescents). The activation in 3-object vs. 1-object trials also significantly increased with age entered as a continuous variable (*p*s < .01), although this effect was less significant than the Age group effect.

3.5.2. Director factor

When comparing the Director Present condition to the Director Absent condition, i.e. when the cues were social stimuli rather than symbols, increased BOLD signal was observed in bilateral superior and middle temporal cortex regions along the superior temporal sulcus and extending into the anterior temporal cortex, as well as in the right inferior frontal gyrus, dorsal MPFC, precuneus, and occipital gyrus (**Table 2** and **Figure 4A**). The reverse contrast, i.e. when the cues were not social but symbolic and rule-based, revealed increased BOLD signal in left parietal cortex only (**Table 2** and **Figure 4A**). There was no significant interaction between Director and Age group factors.

Director × Object interaction

Whole-brain analyses at the cluster FWE-corrected threshold $p < .05$ showed no brain regions with significant Director × Object or Director × Object × Age group interactions. An ROI-approach (which is potentially less robust because it is biased towards particular clusters) was thus used, as follows: mean parameter estimates were calculated for all clusters of the DP > DA contrast and analysed in SPSS using a mixed model repeated measures ANOVA. The ROI clusters were in the left temporal cortex, right temporal cortex, occipital gyrus, dorsal MPFC, precuneus, and right inferior frontal gyrus (see **Table 2** and **Figure 4A**).

Director × Object and Director × Object × Age group interactions were tested using mixed repeated measures ANOVAs on the mean parameter estimates in these ROIs. The aim was to identify regions that showed increased BOLD signal when participants had to take into account the Directors' perspective to perform the task correctly. A significant Director × Object interaction was observed in the left temporal cortex cluster ($F(1, 26) = 6.12$, $p = .020$) and in the dorsal MPFC cluster ($F(1, 26) = 4.87$, $p = .036$). In both cases, the difference in BOLD signal between DP and DA was greater in 3-object than 1-object trials. In addition, the Director × Object × Age group interaction was significant in the dorsal MPFC cluster ($F(1, 26) = 4.76$, $p = .038$). This 3-way interaction reflected the fact that the difference between DP and DA was greater in 3-object than 1-object trials in adults ($p =.014$) but not in adolescents ($p > .9$) (**Figure 4B**). Adolescents thus showed greater BOLD signal in DP than DA in both 1-object and 3-object trials, while adults showed a greater DP than DA activation in 3-object trials specifically, i.e. when the perspective of the directors needed to be taken into account.

Although the Object × Age group interactions of DP vs. DA did not reach significance when comparing the younger and older adolescent sub-groups to the adults ($p = .071$ and $p = .152$ respectively), the same pattern of no difference in DP vs. DA activation between 3-object and 1-object trials (*p*s > .7) was observed in both adolescent group. Similarly, the interaction between Object and age as a continuous variable for DP vs. DA activation ($p = .2$) did not reach significance, suggesting the developmental effect was best accounted for by group (adolescents vs. adults).

3.6. Discussion

This study investigated the development of the neural substrates of action selection when the information guiding the choice is symbolic or social in nature, which enabled us to investigate domain-general processes (common to the symbolic and social conditions) and domain-specific processes (specific to the social or symbolic conditions). The paradigm we used required participants to take the perspective of another person in an implicit manner and respond appropriately in a communicative context. First, we showed that the frontoparietal and temporal brain network showing increased BOLD signal when participants had to identify a target among distractors showed greater BOLD signal in adults than adolescents in the left lateral PFC and parietal cortex (**Figure 3**). Second, the processing of social vs. symbolic stimuli led to social brain network activations in the superior dorsal MPFC, precuneus, and large temporal clusters; the reverse contrast showed activation in the left parietal cortex only, in a region sensitive to the Object factor (**Figure 4A**). Although no whole-brain interaction effects were observed, ROI analyses were performed on those clusters showing greater activations in Director Present than Director Absent trials to investigate specific BOLD signal increases associated with the use of social vs. symbolic cues to guide the selection of the appropriate target among the distractor objects. The left temporal cluster and dorsal MPFC regions showed an interaction between Director and Object, with increased BOLD signal in Director Present 3-object trials, when the perspective of the director needed to be taken into account. The dorsal MPFC region showed a further significant 3-way interaction, between Director, Object and Age group, showing that the adults, but not the adolescents, showed specifically greater Director Present than Director Absent activation in 3-object trials. Adolescents showed greater BOLD signal in the superior dorsal MPFC in Director Present than Director Absent both in 1-object and 3-object trials (**Figure 4**).

3.6.1. Development of the integration of information to guide action selection

The comparison between trials requiring the identification and selection of one of three objects compared to a single target object highlights brain regions recruited in top-down control of attention and goal-directed action. Participants needed to remember the instruction and integrate it with the social or symbolic rule-based cues to identify which of the three exemplar of the target object (e.g. ball) is the correct one to move. All participants were slower and less accurate in 3-object than 1-object trials, but the age groups did not differ (**Figure 2**). Over the whole group of participants, a large bilateral network of brain regions showed greater BOLD signal in 3-object compare to 1-object blocks of trials (**Figure 3A**). Frontal and parietal cortices have been proposed to be the source of spatial attentional modulation of the ventral visual system during object recognition or discrimination (Beck & Kastner, 2009; Corbetta, 1998; Corbetta & Shulman, 1998; Tong, 2003).This network has also been shown to drive non-spatial, feature-based (e.g. colour) selective attention (Giesbrecht, Woldorff, Song, & Mangun, 2003). The network observed in the current study includes the frontoparietal cortex clusters (see **Table 1**) identified as supporting top-down or endogeneous control of selective attention in a spatial cueing paradigm (Hahn et al., 2006), and also those observed in a spatial and feature cueing paradigm (Giesbrecht et al., 2003), and in a visual search paradigm (Booth et al., 2003). Differences in the location of occipital activations and in the spread of activations between these paradigms and the current study may relate to the cueing in the present study being partly based on auditory verbal stimuli, and/or to the greater complexity of the visual stimuli in the present study.

There is little previous research regarding neural changes associated with the development of selective attention. Booth et al. (2003) report greater activations in left thalamus and right anterior cingulate in children (aged 9 -11 years old) compared to adults (aged 20-30) when contrasting a 9 stimuli array visual conjunction search to a simple stimulus detection response; no brain region showed greater activation in adults than in children. In the present study the target stimuli varied on a trial by trial basis, thus BOLD signal changes during the task reflect the encoding and integration of the auditory and visual target information, in addition to the simpler visual detection of the appropriate target shape (e.g. a ball). The developmental results obtained in the current study revealed increased BOLD signal in adults compared to adolescents in the left precentral gyrus extending into the inferior frontal gyrus and in the left intraparietal sulcus (IPS)/supramarginal gyrus when comparing 3-object to 1-object trials (**Figure 3B**). The literature on the development of the neural substrates of attention and cognitive control has not shown a consistent direction of changes in BOLD signal with age (Luna et al., 2010). However lateral PFC and parietal cortex are key regions that consistently show developmental changes. Overall, the pattern of results in the literature suggests that although core regions of the circuitry underlying cognitive control are on-line early in development, the network of brain regions underlying e.g. working memory is still developing during adolescence (Crone & Ridderinkhof, 2010; Luna et al., 2010). The current 3-object vs. 1-object comparison contains a combination of attention (visual search), working memory (remembering the rule, instruction and already attended aspects of the stimuli), and inhibition demands (inhibiting attention towards the distractor stimuli and its associated response). In this context the results show that adolescent development is associated with increasingly bilateral frontal and parietal activations.

In summary the present study shows frontoparietal and temporal cortex activations when participants are required to integrate complex visual and auditory information to search and select one of two possible actions vs. when the action to perform is more simply identified. Despite similar performance, adolescents show hypo-activation of the left frontal and parietal cortex.

3.6.2. Processing of social vs. symbolic information

The instructions were fully matched between the Director Present and Director Absent conditions. However the visual stimuli differed, with two characters standing in front or at the back of the set of shelves in the Director Present condition, vs. the letters F and M and grey or transparent boxes in the Director Absent condition. Main effects of Director were observed for both accuracy and RTs, with better performance in the Director Present condition. Previous research using a behavioural variant of the Director task showed both young and adult participants were much more error prone in the Director Present condition (Apperly et al., 2010; Dumontheil, Apperly, & Blakemore, 2010). However, these studies investigated participants' natural tendency to take into account the director's perspective. In the present study, participants went through a training session where their performance was corrected if they did not take into account the directors' perspectives. This aspect of the task was stressed as important and accordingly we obtained high accuracy rates. The performance benefit associated with the social vs. the symbolic stimuli is in line with previous studies showing that participants perform faster (den Ouden, Frith, Frith, & Blakemore, 2005) and more accurately (Baron-Cohen et al., 1986) on social compared with non-social tasks.

FMRI results showed that the presence of the symbolic stimuli was associated with greater activation in left parietal cortex only, which overlapped with the network of regions more active in the 3-object than 1-object condition (**Figure 4A**). Conversely the presence of the social stimuli was associated with greater activation in a large bilateral network of temporal, precuneus and dorsal MPFC regions that mostly did not overlap with those regions more activated in the 3-object than 1-object contrast, except in the posterior parts of the middle temporal gyrus bilaterally, the precuneus, and part of the left superior occipital gyrus (**Figure 4A**). Thus the presence of social stimuli led to increased BOLD signal in a number of regions that form the social brain network, from face and eye gaze sensitive brain regions along the STS (Haxby, Hoffman, & Gobbini, 2000), body sensitive regions in the extrastriate body area (Taylor, Wiggett, & Downing, 2007) to mentalising regions in the posterior STS and the MPFC (Frith & Frith, 2003; Gallagher & Frith, 2003). More specifically, the STS activation observed in this Director Present versus Director Absent contrast extends into a pSTS region previously observed to show increased BOLD when participants planned or recognised each other's communicative intentions (Noordzij et al., 2010). The MPFC activation was dorsal and aligned with the activations observed in tasks requiring mentalising or trait judgments made on others rather than judgement made on the self (Mitchell et al., 2006; Van Overwalle, 2009). In addition, although the MPFC activation observed in the current study is located in quite a superior dorsal part of the MPFC, recent meta-analyses suggest that mentalising activations extend over a wide range of coordinates in the MPFC (Van Overwalle, 2009, 2011) and that nonstory-based theory of mind studies tend to show more superior activations than story-based theory of mind studies (Mar, 2011). The Director Present condition also likely led participants to associate the male or female voice heard with the male or female director character presented visually. Activations along the

STS have been reported in a study of the encoding of speaker identity in the cortical surface (Formisano, De Martino, Bonte, & Goebel, 2008), and the STS has been shown to represent the integration of auditory-visual integration of faces and voices (Chandrasekaran & Ghazanfar, 2009).

It is noteworthy that the current Director Present versus Director Absent contrast was collapsed across a condition requiring participants to take into account the director's perspective (3-object) and a condition where the director's perspective was not necessary to identify the target object (1-object). Thus, here, the mere presence of the directors and possibly the integration of the auditory instruction and the director's visual representation (Chandrasekaran & Ghazanfar, 2009; Formisano et al., 2008) were sufficient to elicit activations in the mentalising network. This finding is consistent with the suggestion that MPFC plays a broad role in general social cognition (Saxe & Powell, 2006; see also Saxe, Whitfield-Gabrieli, Scholz, & Pelphrey, 2009). An alternative account is that, although mentalising was not necessary in the 1-object condition, participants nonetheless computed the director's perspective. We did not make a distinction between 1-object and 3-object blocks during the training phase or during scanning, thus participants may have computed the director's perspective on all trials rather than deciding whether it was necessary to do so on a block-by-block or trial-by-trial basis. In line with this interpretation, previous research provides evidence that mentalising *can* happen even when it is unnecessary (e.g., Back & Apperly, 2010; Kovacs, Teglas, & Endress, 2010) and even when it actively impedes performance on the main task (e.g., Qureshi, Apperly, & Samson, 2010; Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010).

Our second prediction was that the Director Present vs. Director Absent comparison may show age-related decreases in activation in the dorsal MPFC and increases in the temporal cortex, associated with the processing of social information. No significant Age group × Director interaction was observed, suggesting that the age groups in fact did not differ in their average social brain response to the Director Present stimuli. To summarise, in the present study we observed both domain-general activations in cognitive control regions (3-object vs. 1-object contrast) and domain-specific activations associated with the processing of social cues (Director Present vs. Director Absent contrast).

3.6.3. The use of perspective information to guide action selection

The reliance on social cues was associated with faster reaction times than the use of symbolic cues in 3-object trials (compared with responses in 1-object trials, **Figure 2B**). Whole-brain analyses did not reveal regions exhibiting an interaction between Director and Object factors, and Director, Object and Age groups. However, ROI analyses of the clusters obtained in the Director Present vs. Director Absent contrast showed that the left temporal and superior dorsal MPFC clusters exhibited increased activation in the DP 3-object condition (**Figure 4B**). In the superior dorsal MPFC, this Director × Object interaction was further modulated by Age group, reflecting the fact that the increase in BOLD signal in DP 3-object trials was observed in the adults only (**Figure 4B**). Our third prediction was that age effects on the social brain activation, in particular in the MPFC and temporal cortex, if not observed in the Director Present vs. Director Absent contrast, might have been more specific to the Director Present 3-object condition, which requires the online use of perspective taking information. Our results do show significant age effects in dorsal MPFC (although the pattern in the temporal cortex is qualitatively similar, there was no significant interaction with Age group), however the pattern of changes in activation with age is more complex than predicted from the literature, and discussed below. Note that a limitation of this study is that these findings obtained using ROI analyses are consequently weaker than the whole-brain findings described in the previous sections. Although small, the three-way interaction with age observed in the current study provides interesting new information regarding the development of the social brain during adolescence. Previous studies have consistently reported greater MPFC activations in adolescents than adults in a variety of social cognition tasks (see Blakemore, 2008 for a review of the earlier studies, and Burnett, Bird, Moll, Frith, & Blakemore, 2008; Gunther Moor et al., in press; Pfeifer et al., 2009; Sebastian et al., 2011). The tasks used typically required participants to make an explicit judgement regarding the mental states of a character in a scenario presented in animations (Moriguchi, Ohnishi, Mori, Matsuda, & Komaki, 2007), drawings (Sebastian et al., 2011; Wang, Lee, Sigman, & Dapretto, 2006) or text (Blakemore, den Ouden, Choudhury, & Frith, 2007; Burnett et al., 2008). Two studies required participants to judge how much a phrase (e.g. “I am popular”) described themselves (Pfeifer, Lieberman, & Dapretto, 2007; Pfeifer et al., 2009) and one study involved participants judging a person’s emotion from photos of their eyes (Gunther Moor et al., in press). Thus participants were asked to reflect on their own or someone’s thoughts or emotions in an explicit and somewhat detached manner, and the results showed that in such situations there are greater BOLD signal increases in adolescents than adults in the MPFC. In the present study participants were required to use social cues regarding the perspective and knowledge of another person in an online manner and in a communicative context, and then to perform the appropriate action. Interestingly, our results show that the adolescents did not recruit dorsal MPFC specifically in the perspective taking condition (Director Present 3-object) but more generally whenever the stimuli had a social aspect, i.e. in the comparison of Director Present vs. Director Absent. Adults however, showed greater dorsal MPFC activations that were specific to the Director Present 3-object trials, i.e. when information about the director's perspective had to be taken into account to respond appropriately.

These results cast some light on the possible interpretation of previous findings of greater MPFC activations during mentalising in adolescents. In previous studies it is not clear whether this greater MPFC activation is due to adolescents "over-mentalising" in response to the same stimuli, or having to put in more work in terms of neural resources to achieve the same mentalising computations, or to lower signal to noise ratio associated with increased prefrontal grey matter volumes in adolescence compared to adulthood (see Blakemore, 2008). Recent work using other tasks suggests that decreases in brain activation during adolescence do not necessarily reflect concomitant grey matter volumes decreases (Dumontheil, Hassan, Gilbert, & Blakemore, 2010; Dumontheil, Houlton, Christoff, & Blakemore, 2010). The current study provides no evidence that adolescents use more neural effort to achieve the same mentalising performance: in the absence of differences in performance, adolescents did not show greater activations than adults in Director Present 3-object trials, which require participants to take the director's perspective into account. Instead, adolescents appeared to show less specific mentalising MPFC activations than adults, with MPFC activations observed at a similar level in Director Present 1-object and 3-object trials. These findings are thus more consistent with the over-mentalising interpretation of the greater MPFC activations observed during adolescence in previous studies. The pattern of results observed in the dorsal MPFC in the current study is similar to the finding of a lack of specificity of right TPJ activation in early childhood in a verbal story-based task (Saxe et al., 2009). Among children aged 6-11 years old, right TPJ was similarly recruited when the younger children listened to sections of a story describing a character's thoughts or the physical context, while older children showed right TPJ activation only for mental and not physical facts (Saxe et al., 2009), thus showing increased right TPJ specificity for theory of mind with age.

To summarise, the current study showed that adolescents exhibited dorsal MPFC activation in both social conditions, and did not show the specific increased dorsal MPFC activation observed in adults when the trial required the participant to take into account the director's perspective to choose the appropriate response.

3.6.4. Ecological validity of the task

The communicative nature of the task employed here is more ecologically valid than previous theory of mind tasks employed in neuroimaging experiments. By using this task we are addressing recent concerns that story-based theory of mind tasks might add processing demands that are not directly linked to the computation of mental states (Apperly, Samson, Chiavarino, & Humphreys, 2004). Another advantage of our paradigm is that it contributes to the disentanglement of different theory of mind sub-processes: it is seemingly contradictory that infants know what another person can or cannot see (Flavell, Abrahams Everett, Croft, & Flavell, 1981; Moll & Tomasello, 2006) and that adults calculate other people's perspective automatically (Samson et al., 2010), but that there is a protracted development of performance on our perspective taking task and that even adults have a high error rate. The present study adds to the growing body of work that distinguishes between processes that might be relatively specific to theory of mind, perhaps corresponding to social brain network activations, and processes that are domain-general but equally necessary for actually using theory of mind information to select an appropriate action or verbal response, perhaps corresponding to cognitive control network activations (e.g. Samson, Apperly, Kathirgamanathan, & Humphreys, 2005; Saxe et al., 2004; Saxe et al., 2006; see Apperly, 2011 for discussion). Here, we have shown that both domain-general and domain-specific activations showed developmental changes. Adolescents showed a combination of weaker lateral PFC activations that were not specific to the social condition, and social brain activations, in particular in MPFC, that showed less specificity to perspective taking requirements, compared with adults. The hypo-activation of cognitive control regions in the adolescents may be behind the greater egocentric bias observed in a similar paradigm during adolescence (Dumontheil, Apperly, & Blakemore, 2010). Future studies with greater number of adolescent participants could investigate in more details the observed effects and test their association with pubertal development as opposed to chronological age only (Blakemore, Burnett, Dahl, 2010), as well as the connectivity between the cognitive control and social brain networks.

3.6.5. Conclusion

The aim of this study was to investigate the development of the neural substrates associated with the selection of action among distractors and with the use of social cues to guide action selection. We used a novel paradigm that requires participant to use either symbolic rules or perspective information of other individuals to select an appropriate action in a communicative context. Having shown that the online use of perspective information led to the recruitment of superior dorsal MPFC, left STS and anterior temporal cortex regions in adults, we showed here that adolescents exhibited hypo-activation of domain-general cognitive control regions in the parietal cortex and prefrontal cortex, and hyper-activation of parts of the social brain network, with dorsal MPFC activation observed whether or not the social cues were necessary to perform the action appropriately. These results provide further evidence of the prolonged development of neural substrates of social cognition. They suggest that the pattern of increased MPFC activations in adolescence associated with explicit mentalising judgments (Burnett & Blakemore,

2009) is not found when mentalising has to be used online by taking another's perspective in an active communicative context. Instead, MPFC showed increased activations in adolescence in both the 1- and 3-object conditions, while adults engaged MPFC more for 3-objects. This pattern may reflect over-mentalising in adolescence in conditions where mentalising is not needed (1-object).

In this chapter we have used statistical parametric maps to functionally localize the areas that are activated during our perspective taking task. In the next chapter, we look at the functional integration of the brain responses during the same perspective taking task in adults. For this we will use Dynamic causal modelling to analyse the effective connectivity between the areas that were found to be active according to the SPM analysis.

3.7. Tables

Table 1: Coordinates and Z-values for regions of significant differences in BOLD signal in the main effect contrast of Object [3-object > 1-object], and the interaction between Object and Age group [(Adults 3-object > 1-object) > (Adolescents 3-object > 1-objects)] (*p* < .001 uncorrected at the voxel-level, *p* < .05 FWE corrected at the cluster level).

	L/R	**Brodmann area**	**MNI (x y z)**	**Z-score**	***p*(FWE)**	**Cluster size**
Main effect of Object: 3-object > 1-object						
Superior parietal lobule	R	7	27 -64 52	>8	<0.001	14181
IPS	L	40	-39 -49 46	>8	<0.001	
MTG/occipital gyrus	R	39/19	42 -79 22	>8	<0.001	
SFS	R	6	30 -4 64	>8	<0.001	
Precuneus	R	7	6 -64 49	>8	<0.001	
SFS	L	6	-21 -4 52	>8	<0.001	
IFG	R	44	48 8 34	>8	<0.001	
MFG	R	9	45 26 31	>8	<0.001	
SMG	R	40	39 -40 40	>8	<0.001	
Precentral gyrus	L	6	-42 2 37	>8	<0.001	
ITS	R	37	48 -55 -14	7.61	<0.001	
Occipital gyrus	L	19	-39 -73 -11	6.75	<0.001	
MFG	L	9	-45 23 40	6.32	<0.001	

MFG	L	10/46	-42 47 7	6.22	<0.001	
Medial SFG	R	6	6 14 49	6.20	<0.001	
Thalamus	L		-12 -16 10	4.93	0.010	117
Age group x Object interaction: Adults > Adolescents (3-object > 1-object)						
Precentral gyrus	L	6	-45 5 31	4.96	0.009	368
Inferior frontal gyrus	L	45	-42 23 22	4.60		
Insula	L		-39 17 7	4.12		
IPS/supramarginal gyrus	L	40	-45 -49 43	4.27	0.018	110

IFG: Inferior frontal gyrus; IPS: Intraparietal sulcus; ITS: Inferior temporal sulcus; MFG: Middle frontal gyrus; MTG: Middle temporal gyrus; SFG: Superior frontal gyrus; SFS: Superior frontal sulcus; SMG: Supramarginal gyrus; L/R: Left/Right.

Table 2: Coordinates and Z-values for regions of significant differences in BOLD signal in the main effect contrasts of the Director factor ($p < .001$ uncorrected at the voxel-level, $p < .05$ FWE corrected at the cluster level).

	L/R	Brodmann area	MNI (x y z)	Z-score	*p*(FWE)	Cluster size
Main effect of Director: Director Present > Director Absent						
MTG/STS	L	39	-45 -73 13	7.53	<0.001	1537
STG/STS	L	22	-54 -46 13	5.53	0.001	
MTG	L	21	-48 2 -26	4.91	0.011	
Fusiform gyrus	L	37	-39 -40 -17	4.81	0.017	
MTG/STS	L	21	-54 -10 -14	4.79	0.018	
Hippocampus	L		-33 -10 -20	4.36		
SOG/Cuneus	L	17/18	-12 -94 4	7.15	<0.001	253
MTG/STS	R	22/39	57 -58 13	6.34	<0.001	952
STG/STS	R	22	51 -46 16	5.13	0.004	
MTG	R	21	51 5 -20	4.33		
STG/STS	R	21	48 -31 1	3.9		
MTG	R	21	51 -16 -8	3.87		
Precuneus	L	7	-6 -52 43	5.41	0.001	325
Precuneus	L	7	-12 -52 31	4.59	0.042	
Cingulate gyrus	R	31	6 -55 28	4.3		

Medial SFG	L	8	-12 47 46	4.98	0.008	533
SFG	L	8	-18 38 55	4.84	0.015	
Medial SFG	L	10	-6 56 25	4.06		
Medial SFG	L	10	-9 50 16	3.97		
Medial SFG	R	9	6 53 37	3.94		
SFG	L	8	-21 26 49	3.91		
Cingulate gyrus	L	32	-12 47 4	3.45		
IFG	R	45	57 29 4	4.59	0.042	85
Main effect of Director: Director Absent > Director Present						
Middle occipital gyrus	L	19	-24 -67 37	5.19	<0.001	362
Intraparietal sulcus	L	40	-30 -55 46	4.80		
Supramarginal gyrus	L	7	-42 -55 61	3.94		

IFG: Inferior frontal gyrus; MTG: Middle temporal gyrus; SFG: Superior frontal gyrus; SOG: Superior occipital gyrus; STG: Superior temporal gyrus; STS: Superior Temporal Sulcus; L/R: Left/Right.

3.8. Figures

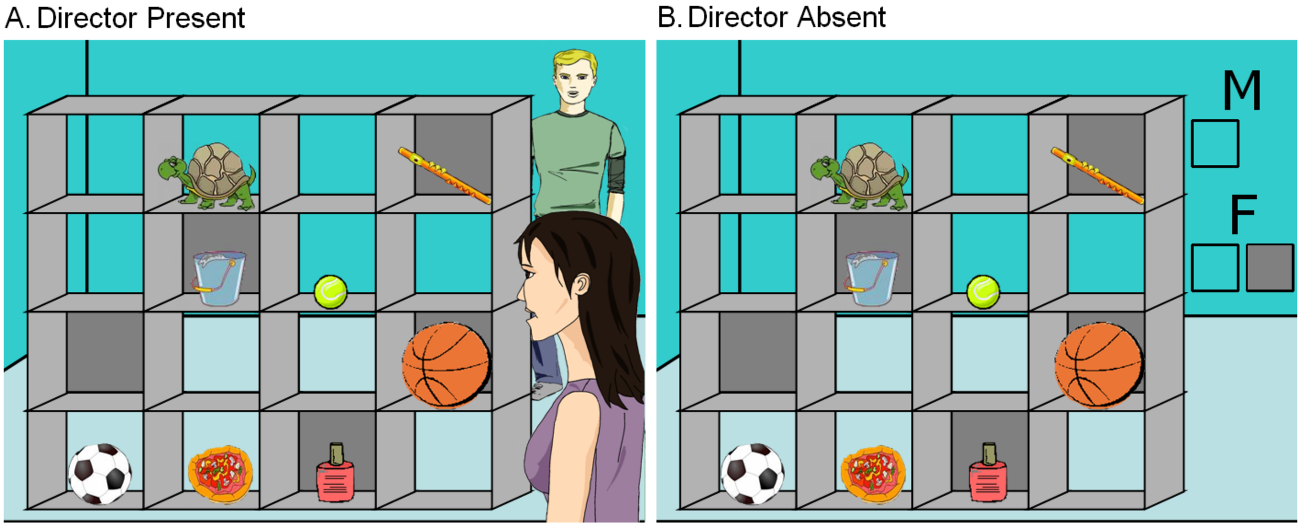


Figure 1: Examples of a 3-object trial in the Director Present (A) and the Director Absent (B) conditions. In both conditions in this example, participants hear the instruction: “Move the large ball up” in either a male or a female voice. In both examples, if the voice is female, the object to be moved would be the basketball, since in the Director Present condition (A) the female Director is standing in front of the shelves and can see all the objects, while in the Director Absent condition (B) the two boxes below the “F” (for “female”) indicate that all objects can be moved by the participant. If the voice is male, the object to be moved would be the football, since in the Director Present condition (A) the male Director is standing behind the shelves and therefore cannot see the larger basketball in the occluded slot, while in the Director Absent condition (B) the single clear box below the “M” (for “male”) indicates that only objects in open shelves can be moved.

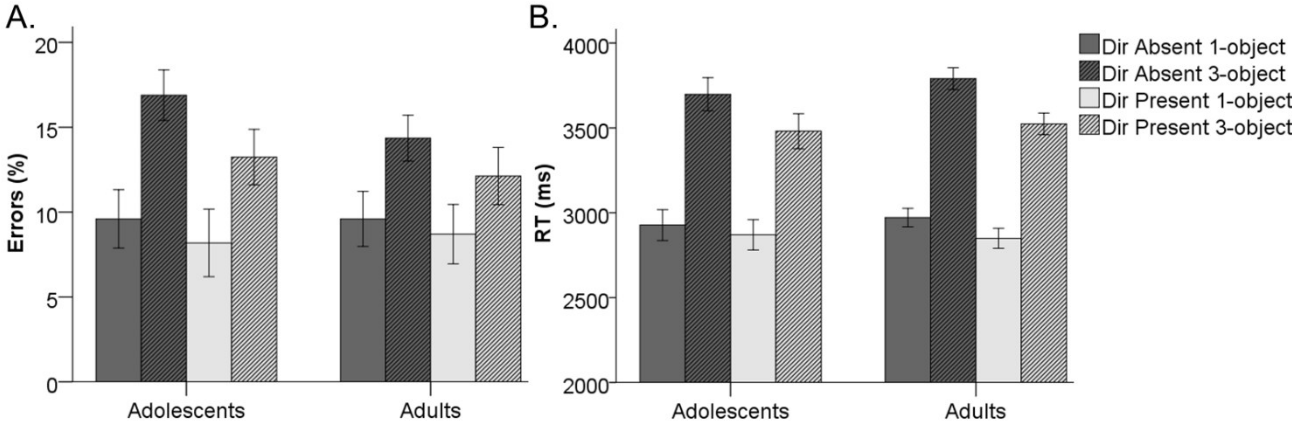


Figure 2: Behavioural performance. (**A**) Percentage errors (mean ± standard error (SE)) in each of the four conditions, plotted separately for the two age groups. There was no main effect of age group, and no interaction between age group and condition. However, overall, participants made more errors in 3-object than 1-object blocks and more errors in the Director Absent than Director Present condition. (**B**) Median RT (mean ± SE) in each of the four task conditions, plotted separately for the two age groups. A similar pattern of effects of the Object and Director factors was observed for RTs as for accuracy. In addition, the interaction between the Object and Director factors was significant, with a larger RT difference between 3-object and 1-object trials in Director Absent (mean 795 ms) than in Director Present (mean 642 ms).

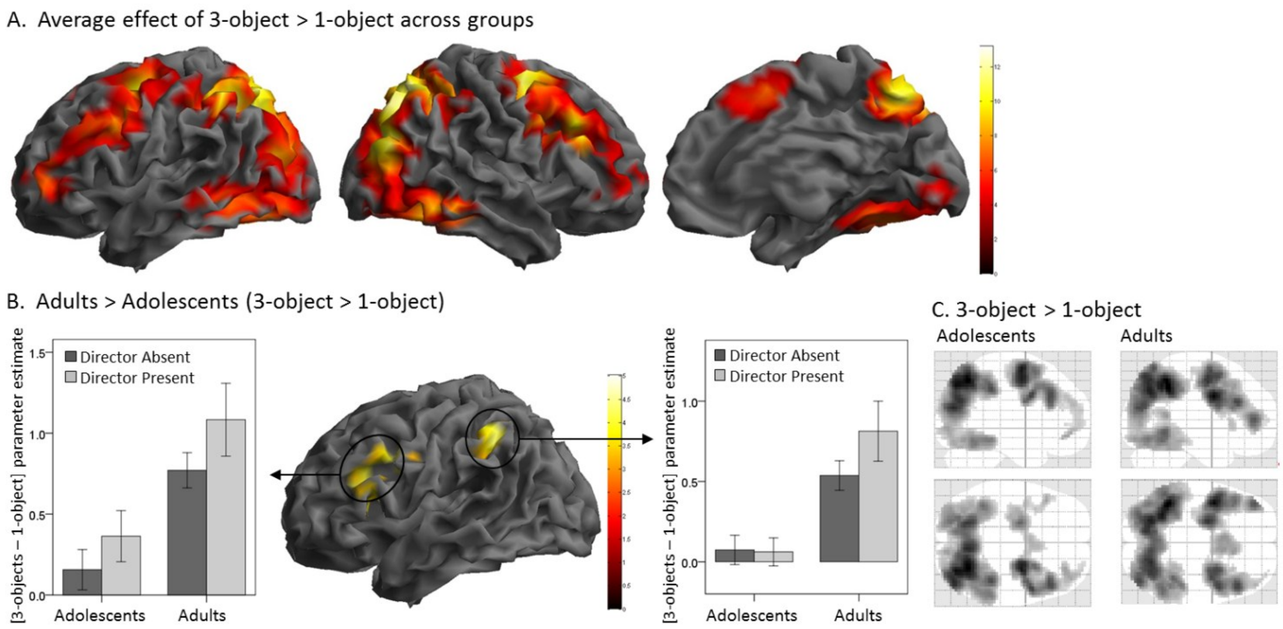

A. Average effect of 3-object > 1-object across groups
B. Adults > Adolescents (3-object > 1-object)
[3-objects – 1-object] parameter estimate
Director Absent
Director Present
1.5
1.0
0.5
0.0
Adolescents
Adults
[3-objects – 1-object] parameter estimate
Director Absent
Director Present
1.0
0.5
0.0
Adolescents
Adults
C. 3-object > 1-object
Adolescents
Adults

Figure 3: Effect of the Object factor and interaction between Object and Age group. (**A**) Main effect of the Object factor. Regions showing increased BOLD signal in 3-object compared to 1-object blocks across age groups are rendered on the SPM8 surface mesh template. From left to right: lateral view of the left hemisphere, lateral and medial views of the right hemisphere. (**B**) Object × Age group interaction. Regions showing increased BOLD signal in 3-object compared to 1-object trials in the adults compared to the adolescents are rendered on the SPM8 surface mesh template (left and right lateral views). Parameter estimates were extracted and plotted for the two significant clusters, in the right IPS (mean of 110 voxels) and the right lateral PFC (mean of 368 voxels). (**C**) On the right, glass brain representations show the regions of increased BOLD signal in the 3-object compared to 1-object blocks in each age group separately. Contrasts were thresholded at $p < .001$ uncorrected at the voxel-level, $p < .05$ FWE corrected at the cluster level.

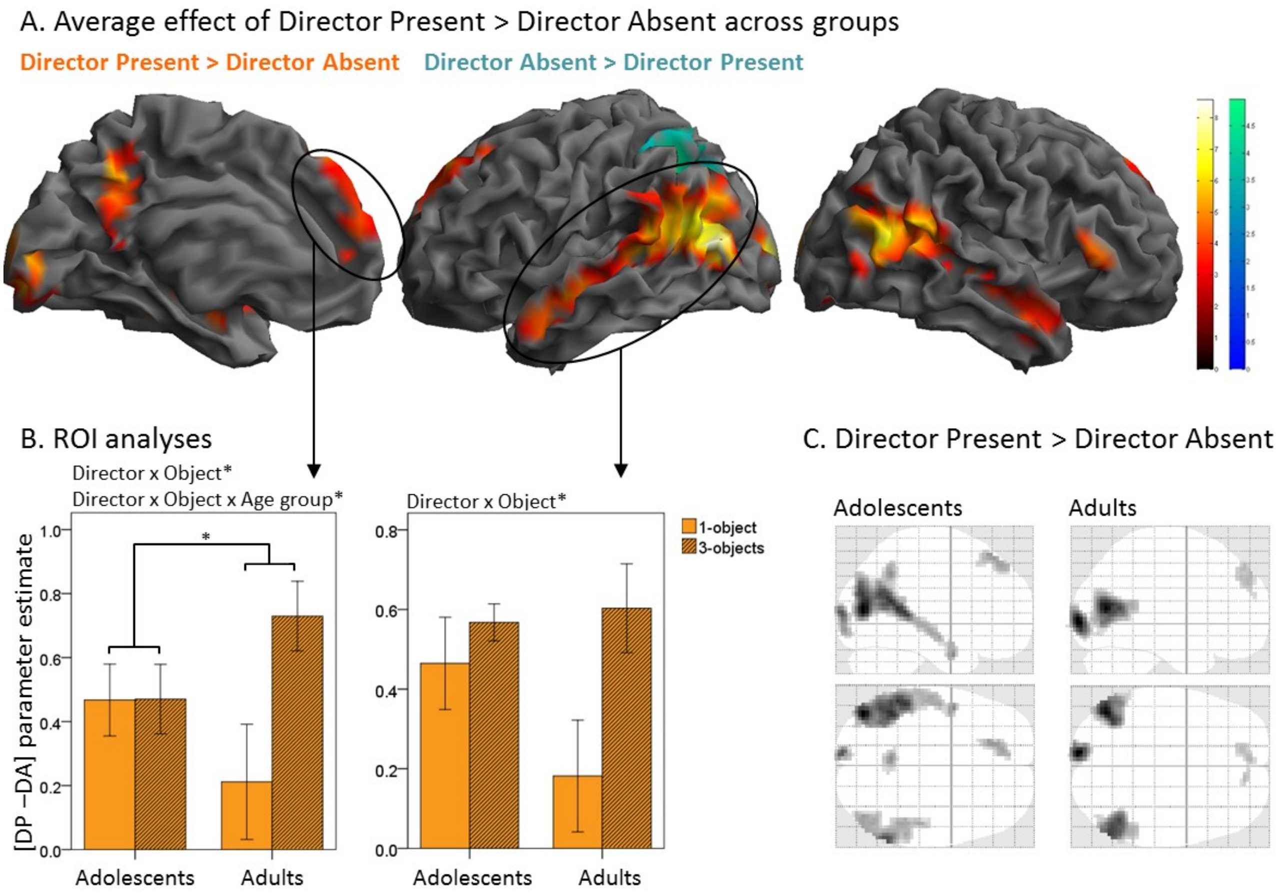


Figure 4: Effect of the Director factor and interaction with Object and Age group. **(A)** Main effect of the Director factor across age groups. Regions showing increased BOLD signal in the Director Present compared to Director Absent blocks (in yellow-red colour scale) or in the reverse Director Absent compared to Director Present blocks (in green-blue colour scale) are rendered on the SPM8 surface mesh template. From left to right: medial and lateral view of the left hemisphere, lateral view of the right hemisphere. **(B)** ROI analyses testing for interactions between Director and Object, and between Director, Object, and Age group. Mean parameter estimates in two clusters of the Director Present > Director Absent contrast across age groups showed a significant interaction between Director and Object factors, with greatest activation in the DP 3-object condition (left temporal cluster, extending along the middle temporal gyrus and

STS, mean of 1537 voxels; superior dorsal MPFC cluster, mean of 533 voxels). The superior dorsal MPFC cluster showed a further significant 3-way interaction between Director, Object, and Age group, with a significant increase in [DP - DA] in 3-object vs. 1-object blocks in the adults only. **(C)** Director Present vs. Director Absent contrast in each age group. Contrasts were thresholded at $p < .001$ uncorrected at the voxel-level, $p < .05$ FWE corrected at the cluster level.

4. Dynamic causal modelling of effective connectivity during perspective taking in a communicative task

This section is partly based on the following manuscript:

Hillebrandt, H., Dumontheil, I., Blakemore, S.-J., & Roiser, J. P. (2013). Dynamic causal modelling of effective connectivity during perspective taking in a communicative task. *NeuroImage*, *76*, 116–124. doi:10.1016/j.neuroimage.2013.02.072

4.1. Introduction

Verbal and non-verbal social interactions both rely on an understanding of other people's mental states, also called Theory of Mind (ToM) or mentalising (Frith and Frith, 2007, 2012; for an excellent review of the extensive research in theory of mind literature we refer the reader to Apperly, 2011). During social interactions, in a complex real-world environment, ToM enables individuals to take decisions and choose actions that are appropriate to the present situation and the inferred mental states of the other people involved. Recent research suggests that it is important to investigate not only ToM development but also how individuals are able to efficiently use ToM information during decision making and reasoning (Samson and Apperly, 2010), and the distinction between ToM-specific processes and executive control (e.g. Saxe et al., 2006; Scholz et al., 2009; Van Overwalle, 2009, 2011; Dumontheil et al., 2012; Meyer et al., 2012).

4.1.1. Studying the neural mechanisms of social cognition with the "Director" task

Keysar and colleagues designed a paradigm to investigate real-world social decision-making, in which participants are faced with a real set of shelves containing objects that are either visible or not visible from the viewpoint of a "director" (a confederate; Keysar et al., 2000, 2003; Lin et al., 2010). The director asks participants to move objects in the shelves and critical instructions require the participant to use information about the director's viewpoint to interpret his instructions correctly. In this Director task, around 50% of the time adult participants fail to use information about the director's perspective and instead erroneously use their own (egocentric) viewpoint when trying to follow instructions (Keysar et al., 2003, 2000). These results were replicated using a computerised version of the task (Apperly et al., 2010; Dumontheil et al., 2010). The Director task differs from other ToM tasks in that it requires participants both to have a functioning ToM to compute the perspective and intentions of another person (the director), and to use this ToM information in concert with other cognitive processes such as attentional and inhibitory control to overcome their egocentric bias and select the appropriate response quickly and accurately (Apperly et al., 2010).

In a previous fMRI study, we employed an adapted version of this Director task (Dumontheil et al., 2010), which in contrast to previous studies that were designed to assess error rates, included extensive task instructions such that participants performed at high levels of accuracy. As in the behavioural version of the task (Keysar et al., 2000), participants followed auditory instructions to move objects in a set of shelves. Using this modified paradigm for fMRI, we found that: (1) selection of an appropriate action when faced with alternatives (Object factor) was associated with domain-general bilateral brain activations located primarily in the frontal and parietal cortices, with additional activations in the inferior temporal cortex; (2) the processing of social information vs. symbolic cues (Director factor) was associated with specific activations in the dorsal medial prefrontal cortex (MPFC) and superior temporal sulcus; (3) the use of social cues as opposed to symbolic cues for the selection of the appropriate action from the alternative options (interaction) was associated with further recruitment of dorsal MPFC and middle temporal gyri, extending into the left temporal pole (Dumontheil et al., 2012, 2010).

Thus, part of the network of brain regions implicated in social cognition, specifically the MPFC and temporal cortex (Brothers, 1990; Frith and Frith, 2007; Van Overwalle, 2009), was recruited when the guiding information was of a social nature compared to more arbitrary symbolic stimuli. Research using visual search paradigms suggests that the prefrontal cortex (PFC) supports the integration of information from the current environment and internal representations, thereby providing a "top-down" influence (i.e. intentionally driven by knowledge, expectations and goals) on attentional orientation and action selection appropriate with current goals (Burgess et al., 2007; Fuster, 2008, 2000; Koechlin and Summerfield, 2007), in contrast with stimulus-driven "bottom-up" mechanisms (Beck and Kastner, 2009; Hahn et al., 2006). Therefore one interpretation of these findings is that the dorsal MPFC, similarly to lateral parts of the PFC, may play a role in providing a top-down influence for the selection of the correct target among distractors when the relevant guiding information is in the social domain. To test this hypothesis we examined the top-down and bottom-up influences of social and executive manipulations on network coupling during the Director task, using Dynamic Causal Modelling (DCM: Friston et al., 2003).

4.1.2. The current study: dynamic causal modelling

DCM estimates the experimental modulation of forwards and backwards connections between regions that are active during a particular task in a directional manner, and thus makes it possible to infer whether experimental manipulations affect top-down or bottom-up influences. We refer to forward and backward connections in the framework of hierarchical predictive coding, in which sensory input is passed forward and processed in the brain hierarchically, from primary sensory to secondary sensory areas, then on to association areas and finally to higher (frontal) areas (Clark, 2012; Friston, 2010, 2005). We used DCM to investigate coupling between frontal, temporal and occipital brain regions (which represented the aforementioned hierarchy in descending order) involved in the Director task, and its modulation by social cues, using fMRI data from a group of adults (Dumontheil et al., 2010).

An important methodological advance in our analysis is the use of a new post hoc model selection procedure (Rosa et al., 2012) to find (1) the best model out of all possible connection architectures with Bayesian model selection (BMS), (2) posterior probabilities resulting from family level inferences testing whether a parameter exists or not, and (3) Bayesian parameter averages (BPA) over all possible models showing how strong fixed connections were and how much they were modulated. Until recently, DCM required very specific hypotheses about the structure of the model (e.g. which connections are modulated by the experimental manipulations). This is because the estimation of each different model takes a few seconds and with increasing number of nodes in each model the combinatorial explosion of possible models that makes it prohibitively expensive in computational terms to estimate all possible models in model space. Instead, we used a new method to find the model evidence for all possible models without estimating them (Friston and Penny, 2011; Friston et al., 2011; Rosa et al., 2012). This approach permits the selection of the winning model as well as family level inferences (Penny et al., 2010) over all possible models to find (1) the probability of certain connections existing and (2) whether these connections are modulated by the experimental manipulations.

We hypothesized that, while occipital and temporal cortex regions process the social aspects of the stimuli in a bottom-up manner (faces and bodies of the directors), the MPFC is involved in the computation, maintenance, and use of perspective information to guide the selection of an appropriate action. These processes are recruited in the Director present vs. Director absent conditions, where the role of the MPFC may be particularly important in the 3-object condition, which requires, on half of the trials, the inhibition of the prepotent bottom-up responses related to one's own perspective.

4.2. Material and methods

4.2.1. Participants

Fourteen adult (mean age 24.9 years, standard deviation (*SD*) 3.0, range 21.3–30.6) right-handed female volunteers included in Dumontheil et al. (2010) were considered for DCM analysis, of which 11 were included in the final analysis (see *Volume of Interest Extraction section below*). All participants spoke English fluently and had no history of psychiatric or neurological disorders. Participants gave written informed consent and the study was approved by the University College London ethics committee.

4.2.2. Experimental design

Our paradigm includes two manipulations embedded in a 2 × 2 factorial design with the factors Director ("Director present" vs. "Director absent") and Object ("3-object" vs. "1-object"). In the Director present conditions, two directors are shown, one female and one male. This enabled the participant to identify easily which director was speaking by the sound of their voice. One director stands behind the shelves, facing the participant, while the other stands on the same side of the shelves as the participant. The position of the male and female directors changed within blocks and was counterbalanced between conditions and within and between participants. Therefore the gender of the directors was not confounded with the different experimental conditions. In the 3-object conditions, the instructions refer to an object that is one of three exemplars in the shelves; the correct object to move depends on which director is speaking and whose perspective to take (see **Figure 1A**). Thus in the Director present 3-object trials, participants need to use the social cues, i.e. the position of the speaking director, to select and move the appropriate object. On half of the Director present 3-object trials the perspective of the director issuing the instruction is different from that of the participant; on the other half the director's and participant's perspectives are the same. This is varied on a trial-by-trial basis, and thus participants need to consider the director's perspective on every trial. Note that this is not an experimental factor (our analyses collapsed across these trial types) but a manipulation that ensures participants integrate trial-specific cues. In Director present 1-object trials, there is no need to take into account the director's perspective to identify the correct object (e.g. "Move the turtle left"), as there are no distractors or other referents; this resembles a bottom-up, visual pop-out as opposed to an effortful top-down visual search (Buschman and Miller, 2007). The Director absent conditions were logically equivalent to the Director present conditions, but the directors were replaced by symbolic cues (see **Figure 1B**).

Stimuli consisted of sets of 4 × 4 shelves with objects located in half of the shelves. Five of the shelves had a grey background (**Figure 1**; see Dumontheil et al. (2010) for details). On each trial, participants were given instructions via headphones, by either a male or a female voice, to move one of the eight objects in the shelves to a different slot, either up, down, left or right (note that this was the participant's left or right). A 2 × 2 factorial within-subject design was used with the factors Director (Present vs. Absent) and Object (1-object vs. 3-object) varying between blocks.

4.2.3. Director factor

In the Director present conditions the display included two directors, one female and one male. In the Director absent conditions, there were no directors in the display (**Figure 1**). Instead, the letter “F” for female and “M” for male were shown beside the shelves. Below each of the letters there was either one transparent box, which indicated to participants that only objects in open shelves should be moved; or two boxes, one grey and one transparent, which indicated that there was no restriction on the participant’s choice and all objects (both in open shelves and occluded shelves) could be moved (see **Figure 1B** for an example). These rules had precisely the same consequences as the position of the director in the Director present conditions. In Director present conditions the physical position of the director issuing the instruction varied on a trial-by-trial basis; similarly, in Director absent conditions the M/F rules changed on a trial-by-trial basis.

4.2.4. Object factor

Instructions in *1-object* conditions (e.g., in **Figure 1**, “Move the turtle left”) referred to a unique target object (there was only one turtle), which was in an open shelf. Instructions in *3-object* conditions (e.g., “Move the large ball up”) could refer to an object in a closed shelf (with a grey background) or an object in an open shelf, which could both be described with the same instruction (e.g. “large ball”). Which of the possible referents was in fact correct was determined by whether the director giving the instruction (identified as male or female by his/her voice) was at the back or front of the shelves (in Director present), or whether the cues indicated that only objects in open shelves could be moved (in Director absent). This manipulation ensured that in Director present 3-object blocks participants had to consider the director’s perspective (which was different from their own perspective on 50% of the trials) in order to know which was the correct object to move. In Director present 1-object blocks the director's perspective made no difference to the correct interpretation of his or her instructions, and thus participants could use their own perspective to select the appropriate object on all trials. In the Director absent condition, perspective taking was not involved. On each trial the visual stimulus and the auditory instruction were presented over a period of 2.2 s, after which the display remained on the screen for another 3.8 s. Four such trials formed one block in our block design and each scanning session consisted of 16 task blocks lasting 24.8 s and four fixation blocks lasting 20 s. Participants responded with their right hand using a trackball mouse (see Dumontheil et al. (2012, 2010); for more detailed methods information).

4.2.5. fMRI data acquisition

3D T_1-weighted fast-field echo structural images and multi-slice T_2-weighted echo-planar volumes with blood-oxygen level dependent (BOLD) contrast (35 axial slices with a voxel resolution of 3 × 3 × 3 mm covering the whole brain; TR = 3 s; TE = 50 ms; TA = 2.9143 s) were obtained using a 1.5 T MRI scanner (Siemens TIM Avanto, Erlangen, Germany). Functional imaging data were acquired in three scanning sessions (one session of one participant was discarded due to low accuracy) lasting approximately 8 min 40 s each in which 174 volumes were obtained. The first 2 volumes of each session were discarded to allow for T_1 equilibrium effects. A T_{1}-weighted anatomical image lasting 5 min 30 s was acquired after the first two functional sessions for each participant.

4.3. Data analysis

4.3.1. fMRI data

fMRI data were analysed using Statistical Parametric Mapping (SPM8 for the GLM and SPM12a for DCM, www.fil.ion.ucl.ac.uk/spm8). Detailed methods of the preprocessing and General Linear Model (GLM) analysis of the fMRI data can be found in Dumontheil et al. (2012). Here we report results of data reanalyzed with concatenated sessions for the purpose of DCM. Thus, the results of the GLM differ slightly due to the concatenation of sessions. Briefly, images were realigned, slice timing corrected, and normalised to a standard EPI template based on the Montreal Neurological Institute (MNI) reference brain. The resulting 3 × 3 × 3 mm images were spatially smoothed with an 8-mm Gaussian kernel. The time series were modelled with boxcar regressors of the instructions, fixation, four types of task blocks, and with one event-related regressor representing error trials. Furthermore, we included constant session effects. Appropriate stimulus functions were convolved with the canonical hemodynamic response function to form regressors. Together with regressors representing residual movement-related artifacts and the mean over scans, these regressors comprised the full model for each session. A flexible factorial 2nd level analysis was performed to identify significant regional effects for the Director present – Director absent contrast and the Director × Object interaction contrast ([Director present 3-object – Director present 1-object] – [Director absent 3-object – Director absent 1-object], masked inclusively by [Director present 3-object – Director present 1-object]) to ensure that any interactions observed were driven by effects in the Director present condition (see **Table 1**).

4.3.2. Dynamic Causal Modelling

DCM is a Bayesian framework for modelling and inferring the directed connectivity among hidden (unobserved) neuronal states from measurements of brain activity, in this case Blood

Oxygenation Level Dependent (BOLD) activity. It can be used to analyse task or set-dependent effective connectivity i.e. changes in coupling strength, providing information about the changes in directed influence of one area over another in certain psychological contexts. In other words, DCM can be used to determine how activity in higher brain regions is caused (we speak of causality here in the context of control theory, see Marreiros, Kiebel, & Friston, 2008) by activity in lower sensory areas and vice versa. The generative model used by DCM is based on coupled bilinear differential state equations modelling distributed brain activity and canonical haemodynamics for each region (Friston, Harrison, & Penny, 2003).

4.3.3. Volume of Interest Extraction

Two volumes of interest (VOIs) were extracted based on the peaks of the Director present > Director absent contrast (see Results section for GLM results and Dumontheil et al. 2010). The first VOI was located in the left superior occipital gyrus (SOG) (-12 -94 4, note that this peak coordinate was labelled as left cuneus in Dumontheil et al., 2010, because the activation was more diffuse due to a less conservative statistical threshold). The domain general function of the SOG is visual perception relevant for motor control (Lui et al., 2006). Activation in this region is observed in more specific social contexts such as during the perception of bodies (Kret et al., 2011). The second VOI was the left middle temporal gyrus (MTG) (-45 -70 13). The MTG domain general function is thought to be motion processing and multisensory integration (Hein and Knight, 2011; Onitsuka et al., 2004), and is recruited in more specific social cognitive tasks during face and speech processing (for a review see, Hein and Knight, 2011). A third VOI was extracted based on the most active peak from the (masked) interaction contrast. This VOI was located in the left dorsal MPFC (-9 38 34). The MPFC is thought to support the processing of one's own and others' mental states (Amodio and Frith, 2006; Van Overwalle, 2009).

Timeseries from VOIs associated with different contrasts were summarised using the SPM12 Eigenvariate toolbox: we extracted each participant's principal eigenvariate around the participant-specific local maxima activation nearest to the peak voxel of the group (2nd level) GLM analysis (see **Table 1**). All VOIs were taken from the left hemisphere, which showed higher levels of activity, possibly due to the language demands of the task, the fact that participants responded using a trackball mouse held in their right hand and/or the fact that the directors were presented in the right visual field. For the purpose of VOI extraction the normalised images were smoothed with a 4 mm Gaussian kernel to improve anatomical accuracy. The radius of the VOI spheres was 8 mm and the search radius for local maxima from the group analysis was restricted to 16 mm for the main effect and 20 mm for the interaction, which had a more diffuse activation. The rationale for using a more liberal radius for the interaction pragmatic: when using the same radii, we would not have been able to extract significant voxels from too many participants. The justification for this was twofold: first, a statistical test for an interaction effect is less powerful than a test for main effects and second, activity in prefrontal regions is less robust and inter individual variability is higher. All voxels contributing to the eigenvariates were significant at $p < 0.05$ uncorrected and adjusted at $p < 0.05$, for the effects of interest (i.e. only for those regressors that were used in the DCMs for input or modulation). We were unable to create VOIs for 3 participants from the interaction contrast, as they did not show any activation above threshold within the search radius. The interaction contrast had less power (due to a lower number of trials in each level contrast) and showed weaker effects overall than the main effect. Hence, we conducted all analyses on 11 of the initial 14 participants.

4.3.4. Specification of dynamic causal models

We created and estimated DCMs (Friston et al., 2003) with DCM12 (version 4750) as implemented in SPM12a. The DCM estimation routines in SPM12a differ slightly from those in SPM8 and DCM10, in that the hyperpriors have been adjusted to reflect a more realistic signal-to-noise ratio in regional (VOI) timeseries. The DCMs were based on the VOIs reported above and used the main effects of director and object to modulate the connections between regions. All our DCMs were deterministic (as opposed to stochastic, see Daunizeau et al., 2012), bilinear (as opposed to nonlinear, see Stephan et al., 2008), two-state models (Marreiros et al., 2008), with mean-centred inputs. Two-state DCMs differ from one-state models in that activity in one brain region is modelled so that is has both an excitatory and an inhibitory neuronal population, and introduces positivity constraints that allow extrinsic (between regional) populations influences of one region on another to only be excitatory (Marreiros et al., 2008). To simplify the models and for ease of interpretation, we disallowed the modulation of self-connections within each region. In the present study the diagonals in DCM.B matrix, which describes the change in coupling strength (Friston et al., 2003), were thus set to zero, while all other connections were set to one. Self-connections were only present in the DCM.A matrix, which represents the *fixed* connection strength between areas (also referred to as endogenous, direct, context-independent, or average connectivity; see Friston et al., 2003). The driving input into the model – represented by the DCM.C matrix (Friston et al., 2003) – was the main effect of Director present vs. absent. This driving input entered the most posterior region, the SOG. Our hypothesis was that this region would be the first region showing sensitivity to the presence of cues (faces and bodies), indicating the Director present vs. absent, and that it would subsequently influence activity in more anterior regions.

4.3.5. Post-hoc Bayesian model selection

Until recently, one aspect of DCM was that one had to have specific hypotheses about the structure of the model (e.g. which connections are modulated by the experimental manipulation, Stephan et al., 2010). A model space with n nodes has $2^{n \times n}$ permutations of connections that can be turned on or off, which can be modulated by different experimental manipulations (Friston and Penny, 2011). This combinatorial explosion makes it prohibitively expensive to estimate a large number of models (Friston et al., 2007). Instead, we used a new method to find the model evidence for all possible models by only inverting (estimating) the full model (Friston & Penny, 2011; Rosa et al., 2012) to select the winning model. This was achieved by searching over all possible models to find the probability of certain connections existing and estimating whether certain connections are modulated by certain experimental manipulations or not (Friston & Penny, 2011; Rosa et al., 2012).

This approach fits the full model – with all free parameters – to the data. The full model generally contains all possible intrinsic forward and backward connections, and all inputs and modulations of these connections by experimental factors. One then approximates the evidence for all possible reduced models, which have fewer parameters and are therefore nested within the full model. This is achieved by setting the prior variance over all combinations of free parameters (to zero). Based on the posterior density over the parameters of the full model, the approximate evidence for each reduced model can then be obtained (Friston & Penny, 2011; Rosa et al., 2012). These post-hoc estimates of model evidence and the (conventional) free energy approximation (following inversion of reduced models) have been shown to yield very similar results with both simulated and real data (Friston & Penny, 2011; Rosa et al., 2012). This method assumes that that the prior density of both the mean and covariance are Gaussian (Friston

& Penny, 2011; Rosa et al., 2012). This approach has recently been criticised (Lohmann, Erfurth, Müller, & Turner, 2012; but see, Friston, Daunizeau, & Stephan, 2013).

Below, we first present the results of Bayesian model selection comparing all models at once to find the winning model. Then we present the model posterior probability over parameters (with and without a given free parameter) of whether a fixed connection or a particular modulation exists at all using family-level inferences (Penny et al., 2010). The posterior probability is the probability that a model (or family of models) provides the best explanation for the measured data across participants (Penny et al., 2004). The log-evidences for all subsequent analysis were pooled in a fixed effects fashion, because we assumed that the underlying model structure did not vary across the participants for whom VOI timeseries could be extracted. Then we present the average BPA parameter estimates for the model with the highest evidence (the winning model) to elucidate the quantitative nature of the connection e.g. how much a connection is modulated or how much fixed connectivity there is (Friston et al., 2003). BPA computes a joint posterior probability density over parameter estimates for a group of participants, by using the posterior from one participant as the prior for the next participant, whose posterior then serves as the prior for the next participant etc. (Kasess et al., 2010; Stephan et al., 2010). Note that the fixed and modulatory parameters were always scale parameters (exponentiated) to ensure positivity as per convention for two-state DCMs, so that the extrinsic connections are always excitatory (Marreiros et al., 2008). Scale parameters of two-state DCMs are thus higher than parameter estimates from one-state DCMs. Our unexponentiated parameter estimates ranged from -1.3 to 1.8 Hz, similar to one-state DCM parameter estimates reported in other studies (Goulden et al., 2012; Rosa et al., 2012).

In addition to the average BPA estimates, the post-hoc optimisation also provides BPA parameter estimates for individual participants that can be compared with conventional frequentist statistics (Stephan et al., 2010). We supplemented our Bayesian inference with analysis of variance of individual changes in connection strengths to demonstrate the consistency of these effects across subjects.

4.4. Results

4.4.1. General Linear Model results

As reported previously (Dumontheil et al., 2010), the Director present > Director absent contrast, performed with a whole-brain voxel-level FWE-corrected threshold of $p < .05$, showed increased BOLD signal in the left middle temporal gyrus (MTG VOI), left superior occipital gyrus (SOG VOI), and right middle temporal gyrus (**Table 1**). We performed three small volume corrections (FWE $p < .05$) in the left posterior superior temporal sulcus (STS)/temporo-parietal junction (TPJ), left temporal pole and left MPFC on the basis of coordinates from previous studies (Dumontheil et al., 2010, for details). This revealed significant interactions in the left middle temporal gyrus and left dorsal MPFC (**Table 1**); the latter was the most significant after small volume correction. Hence, this region was selected as a VOI for the DCM analyses.

4.4.2. Dynamic Causal Modelling results

DCMs were created using the three VOIs described above: SOG; MTG; and MPFC in the left hemisphere (**Table 1**). The post-hoc analysis (see Material and Methods for details) finds the best model and also furnishes the posterior probability of whether individual parameters exist or not. The latter is equivalent to family comparison inferences, which test whether a family of models without a certain parameter (e.g. a connection between two areas) has a higher probability than the family of models with this parameter (Penny et al., 2010). Finally, we show the results of the parameter estimates using the BPA of the winning model's parameter estimates.

4.4.3. Bayesian model selection and family-level inferences

We first assessed the model with the best evidence (a metric in which model fit is traded off against model complexity). A comparison of the evidence for all possible models showed that the winning (optimal) model with the highest probability had a probability of 0.30 (**Figure 2A**). The winning model was the full model that had all connections and all modulations **(Figure 2B)**. The next most probable model's probability was 0.27. When comparing this model to the winning model, the Bayes factor is only 1.11. In Bayesian statistics, this would not be considered strong evidence for the full model over the next probable model. In frequentist, classical hypothesis testing this would correspond to a non-significant difference. However, traditional Bayesian as well as frequentist approaches are used for smaller model spaces. In large model spaces like ours it is to be expected that the posterior mass is diluted over a high number of models and no single model has a very high probability (Rosa, Friston, & Penny, 2012), since all models that share some characteristics with the full winning model will have a non-zero probability in the presence of any noise. For instance, using the post-hoc approach with synthetic data, Rosa et al. (2012) were able to obtain the model they knew a priori to be the true model in a model space that was even smaller than ours (half the size) and a Signal to Noise Ratio (SNR) of 2.6. However, when they compared this winning true model (that the data was generated from) to the next probable (false) model the resulting Bayes factor was just 1.94 (which is comparable to ours). **Figure 2A** shows the (log-)posterior probability of all models examined (256 models overall). Crucially, the ‘‘best model’’ approach used in Bayesian model selection is very useful but can become brittle (as in software brittleness; see Penny et al. 2010), when one compares a large number of models (Penny et al. 2010) so, in addition, we performed a family level inference.

When assessing the probabilities for each parameter existing or not using family-level comparison (Penny et al., 2010), we found evidence for reciprocal fixed connectivity between all three regions and modulation of all connections. The results of the family-level inference of all models showed that the posterior probability (over parameters) for reciprocal fixed connectivity between all the VOIs was (almost) 1. Moreover, the evidence that these connections were all modulated by both contrasts (main effects of director and object) was also (almost) 1.

Since the full model had the highest evidence, this indicates that none of the alternative, reduced models outperformed the full model. Moreover, because the family-level inferences also showed that all fixed connections were modulated by both modulators, the full model was selected for further analyses of the parameter estimates in order to examine the strengths of the fixed connectivity and the strengths of the modulation by experimental manipulation.

4.4.4. Comparison of connection strength

We tested whether there were quantitative differences between the selected model's Bayesian parameter estimates to find out whether connections between different areas were subject to different modulation. To achieve this, we conducted repeated measures ANOVAs with SPSS 20.0 for both modulators (Director present vs. Director absent and 3-object vs. 1-object) on the individual participants' (BPA) parameter estimate values of the optimal model with connection as a factor with 6 levels. While the two state-DCMs use exponentiated scale parameters that introduce positivity constrains and are so more plausible to interpret, these values are likely to not be normally distributed and heteroscedastic, because the exponential function is the inverse function of the natural logarithm (which is commonly used to transform data to meet the assumption of a normal distribution, see (Bland and Altman, 1996a, 1996b). Thus, we used the original unexponentiated non-scale parameter estimates for these statistics (see **Table 2**), but plot the exponentiated values in **Figure 3 and 4A-B**. When significant, the main effect of connection was followed up by paired post-hoc comparisons between the connections.

The 6 (connections) × 2 (modulator) repeated measures ANOVA showed no significant main effects of connection or modulator, but a significant connection × modulator interaction ($F(5,50) = 8.26, p < 0.001$). When considering the modulators separately, there was a main effect of connection for both the Director present vs. absent modulation ($F(5,50)=4.26, p=0.003$), and the 3-object vs. 1-object modulation ($F(2.4,24.4)=8.93,p=0.001$; Greenhouse-Geisser corrected). Post-hoc tests showed that connections from the MPFC to the SOG and MTG, and the lateral reciprocal connections between them, were more strongly modulated by the Director present vs. absent modulation (**Table 2;** only post-hoc comparisons of connections with connections going to the MPFC are displayed in **Table 2**) than were the connections to the MPFC. By contrast, the 3-object vs. 1-object manipulation modulated connections leading to the MPFC more strongly than those projecting from this region or the reciprocal connections between the SOG and MTG (**Table 2**). We did not control for multiple comparisons when analysing the post-hoc comparisons of connections, because there was no selection process: we analysed and reported all possible connections and they were all at least marginally significant.

4.5. Discussion

We used DCM to explore effective connectivity between SOG, the MTG, and the MPFC, in a task that required participants to take into account another person’s perspective in order to guide action selection. We used a novel post-hoc model selection routine (Friston and Penny, 2011; Rosa et al., 2012) to look at all possible dynamic causal models. Family-level inference provided strong evidence for reciprocal fixed connectivity between all three areas and modulations by both the Director and the Object factors, which respectively manipulate the nature of the stimuli (social vs. non-social) and the need for top-down influences on action selection.

4.5.1. Fixed and modulatory connectivity

Our family-level inferences suggest that there is high probability of reciprocal fixed connectivity between all three areas that was context-independent and thus driven by the input of the Director present vs. absent manipulation. Our hypothesis was that, similar to the role played by lateral parts of the PFC during tasks that are novel or not automatic and require attentional control and action selection (Burgess et al., 2007; Fuster, 2008, 2000; Koechlin and Summerfield, 2007), the MPFC plays a role in the control of attention and action selection when the guiding and informative cues are of a social nature. We thus predicted that, in the condition that requires perspective-taking information to be used to identify the correct target object (Director present 3-object), we would observe increased effective connectivity from the MPFC towards the SOG and MTG. We found strong evidence that all connections were modulated by both the social and executive manipulations – however, we were also able to disambiguate the relative contribution of modulatory strength on different connections by the two different experimental manipulations.

4.5.2. Social manipulation

Consistent with our prediction, analysis of the modulatory effects revealed that the presence of a social cue (Director present vs. absent contrast) increased the strength of the backward connections from the MPFC more strongly than the forward connections from the SOG and the MTG. Interestingly, this was also the case for the backward connection from the MTG to the

SOG. The dorsal MPFC consistently shows increased activation in response to social stimuli and mentalising, in particular when participants are thinking about others' thoughts, intentions or personality traits, and has been suggested to hold the representation of other people's mental states (Tamir and Mitchell, 2010). The left MTG region was close to TPJ which is also consistently implicated in social cognition and perspective taking (Frith and Frith, 2006). Our interpretation of these findings is that the social context leads the dorsal MPFC and the MTG to increase activation in regions processing socially-relevant information, and also to increase the coupling between these regions.

4.5.3. Attentional and executive control manipulation

The 3-object vs. 1-object factor increased bottom-up connectivity by modulating the forward connections of the SOG and the MTG to the MPFC more strongly than the backward connections. This may reflect the transmission of the socially relevant information (location of the directors, presence or not of the grey background on the two possible target objects) to the MPFC for further computation of viewpoint and intention of the relevant director. In 3-object trials, the position of the directors, and the relative position or relative size of the three objects needed to be considered to infer the intention of the director correctly. This was associated with longer reaction times (see Dumontheil et al., 2010) and increased demands of processing both the social and non-social information. Note, that in previous perspective-taking tasks (Aichhorn et al., 2006; David et al., 2006; Vogeley et al., 2004) the participants only judge whether another person can see a particular object, whereas in our task, participants have to infer and select which object another person refers to, based on their perspective (Dumontheil et al., 2010). However, executive demands sometimes increase activity in MPFC (Levy and Wagner, 2011) areas and social demands increase activity in MTG and SOG (as evidenced by our fMRI data), which may at least partly account the corresponding modulating effects on the connectivity between these brain areas. Furthermore, the visual search, and the computation and selection of the correct action in the 3-object vs. 1-object condition, which we have summarized as increased executive demands like the inhibition of prepotent responses and working memory demands, is likely to have also required more attention and cognitive resources (Levy and Wagner, 2011). This is partly supported by the activation of regions associated with attention and control in the 3-object vs. 1-object main effect contrast (Dumontheil et al., 2012, 2010; Duncan, 2010). Finally, since SOG and MTG are essentially sensory areas, the top-down predictions to these areas can only serve to mediate the inhibition of prepotent responses by selectively reducing the attentional capture by incorrect but prepotent stimuli (Ji and Neugebauer, 2012; Levy and Wagner, 2011; Sharp et al., 2010).

4.5.4. Bayesian hierarchical predictive coding during social behaviour

Our results can be interpreted within the framework of recent theories that attempt to explain (social) behaviour in terms of Bayesian hierarchical predictive coding (Brown and Brüne, 2012; Clark, 2012; Friston, 2010). Predictive coding postulates that forward, bottom-up connections pass prediction error signals with (unexpected) sensory information about the stimulus from ‘lower’ (sensory) areas to areas ‘higher’ in the cortical hierarchy (Friston, 2010). In this framework, the top-down backward connections pass predictions based on an internal (generative) model about the stimulus to lower sensory areas to minimize sensory prediction error (by selectively sampling the stimulus array) and to induce behavioural responses (Clark, 2012; Friston, 2010). In our 3-object factor, the stimulus array has to be sampled more extensively than in the 1-object condition like in traditional top-down visual search paradigms (see for instance, Buschman and Miller, 2007) to find the correct target (‘which ball is referred to?) amidst distractors, and so ‘surprising’ unpredicted sensory information about distractors is passed forward in the cortical hierarchy as prediction error, unlike in bottom-up searches (where a ball ‘pops out’ among trucks). Lastly, the main effect of director not only modulates particularly strongly the backward connections (potentially by passing down predictions about the director), but also modulates what is usually characterized as a forward connection from SOG to MTG (**Figure 4A**). Because the task requires multisensory integration of visual stimuli with auditory instructions, the connection from SOG to MTG might be a backward connection where the SOG predicts auditory responses in the MTG in a top-down fashion.

4.5.5. Limitations

Our study has some limitations. First, by using the DCM post-hoc procedure, like Rosa et al. (2012), our posterior mass was diluted over a large number of models and so we were not able to reach a statistically significant Bayes factor when comparing our winning model with the next probable model in our large model space (see results section for more details). This is important to consider for future research using the post-hoc estimation routines with large model spaces. Furthermore, we did not test for any potential nonlinear modulatory gating effects that the activity in brain regions exerts on the connections between them (Stephan et al., 2008) in order to simplify the model and ease interpretability. For the same reason, we chose to look only at brain connectivity in the left hemisphere and omitted extracting a VOI around the right MTG, which was active during Director present vs. Director absent contrast. Indeed, as we only included three left hemisphere regions recruited in the social (Director present) condition, and not right hemisphere regions or regions responsive to the executive demands, our more constrained model would have failed to capture mediating effects and connections between hemispheres and between social and executive regions enabling task performance.

4.5.6. Summary

Our results suggest that effortful search and selection of the correct mental state of other people for the purpose of action selection involves lower sensory areas and multisensory integration areas (SOG, MTG, Hein and Knight, 2011) as well as prefrontal areas (MPFC). We provide evidence for an account in which social predictions based on an internal model modulate backward connections, whereas when a perspective diverges from one's own perspective one has to search for the correct perspective, which modulates forward connections.

In order to take the perspective of another person one first has to recognise an entity as animate. The neural mechanisms underlying perception of animate motion have not been explored. In the next chapter, we use DCM to investigate the effective connectivity underlying animacy perception.

4.6. Tables

Table 1: GLM results.

The Director present > Director absent contrast was performed with FWE correction at $p < 0.05$, whole brain level. For the Director × Object [(Director present 3-object > 1-object) – (Director absent 3-object > 1-object), inclusively masked by (Director present 3-object > 1 –object)], only those regions are listed that survived small volume FWE correction at $p < 0.05$ based on a 12mm sphere centred on coordinates from previous literature are listed (see Dumontheil et al., 2010 for details). * and $^{\#}$ indicate FWE-corrected values for the whole brain or a small volume (12 mm sphere) respectively.

Area	**MNI Coordinates**	**Cluster size**	**Z-value**	**p-value for peak coordinate (FWE-corrected)**
Director present vs. Director absent				
Left middle temporal gyrus	-45 -70 13	16	5.07*	0.006*
Right middle temporal gyrus	60 -64 13	6	4.95*	0.01*
Left superior occipital gyrus	-12 -94 4	9	4.91*	0.012*
Director × Object interaction				

Left superior dorsal medial prefrontal cortex	-9 41 34	5	3.96[#]	0.004[#]
Left middle temporal gyrus	-36 8 -26	4	3.33[#]	0.027[#]

Table 2: Comparison of modulatory connection strengths.

Post-hoc paired comparisons of connections strengths of the modulatory effects of the Director present > absent manipulation and the 3-object > 1-object manipulation. Connections leading to the SOG and MTG were significantly more strongly modulated by the Director present vs. absent manipulation than those leading to the MPFC. By contrast, the 3-object vs. 1-object manipulation modulated the connections leading to the MPFC significantly more strongly than those leading to the MTG and SOG. For clarity, we only report the values of regions going to the SOG and MTG for the Director present vs. absent contrast and the values going to the MPFC for the 3-object vs. 1-object here.

Modulation by Director present vs. absent contrast

Backward/lateral connection	Forward connection	Mean Difference between backward/lateral - forward connections	Std. Error	*p* value
MTG to SOG	SOG to MPFC	0.446*	0.121	0.004
	MTG to MPFC	0.530*	0.163	0.009
MPFC to SOG	SOG to MPFC	0.488	0.267	0.097
	MTG to MPFC	0.572*	0.179	0.009
SOG to MTG	SOG to MPFC	0.507*	0.122	0.002
	MTG to MPFC	0.591*	0.209	0.018
MPFC to MTG	SOG to MPFC	0.456	0.214	0.059
	MTG to MPFC	0.540*	0.151	0.005

Modulation by 3-object vs. 1-object contrast

Forward connection	Backward/lateral connection	Mean Difference between forward - backward/lateral connections	Std. Error	*p* value
SOG to MPFC	MTG to SOG	0.759*	0.271	0.019
	MPFC to SOG	1.091*	0.271	0.002
	SOG to MTG	0.896*	0.211	0.002
	MPFC to MTG	0.867*	0.203	0.002
MTG to MPFC	MTG to SOG	0.473*	0.18	0.025
	MPFC to SOG	0.804*	0.151	<0.001
	SOG to MTG	.609*	0.224	0.022
	MPFC to MTG	.581*	0.162	0.005

*. The mean difference is significant at the .05 level.

4.7. Figures

Figure 1 (color for the Web)

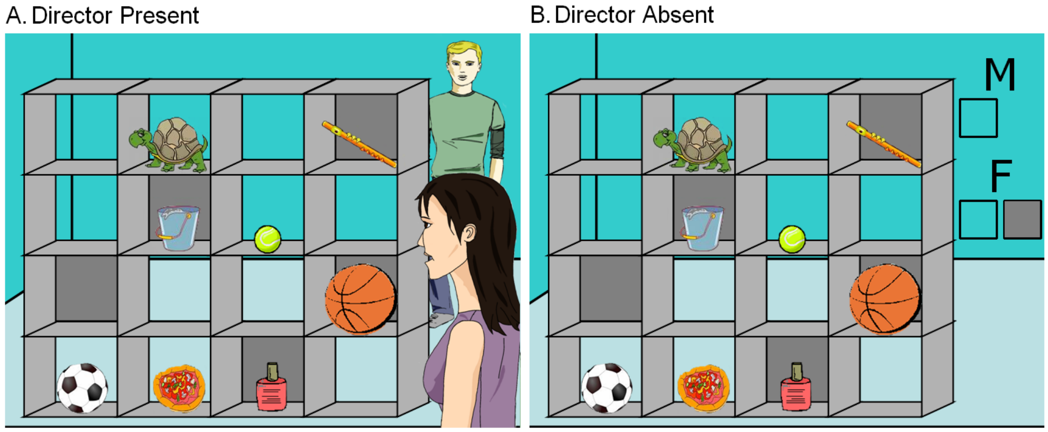


Figure 1: Examples of a 3-object trial in the Director present (A) and the Director absent (B) conditions. In this example, participants hear the instruction: “Move the large ball up” in either a male or a female voice. Two of the objects could correspond to this instruction depending on the director’s viewpoint. The largest ball (or equivalent) was always located in a closed shelf (not visible from the back), while the second largest ball was located in an open shelf. Here, if the voice is female, the object to be moved would be the basketball, since in the Director present condition (A) the female Director is standing in front of the shelves and can see all the objects, while in the Director absent condition (B) the two boxes below the “F” (for “female”) indicate that all objects can be moved by the participant. If the voice is male, the only appropriate response would be to move the football, since in the Director present condition (A) the male Director is standing behind the shelves and therefore cannot see the larger basketball (which was thus a distractor), while in the Director absent condition (B) the single a transparent box below the “M” (for “male”) indicates that only objects in clear slots can be moved by the participant.

The third object was another type of distractor that did not fit any of the perspectives (the tennis ball).

Figure 2 (color for the Web)

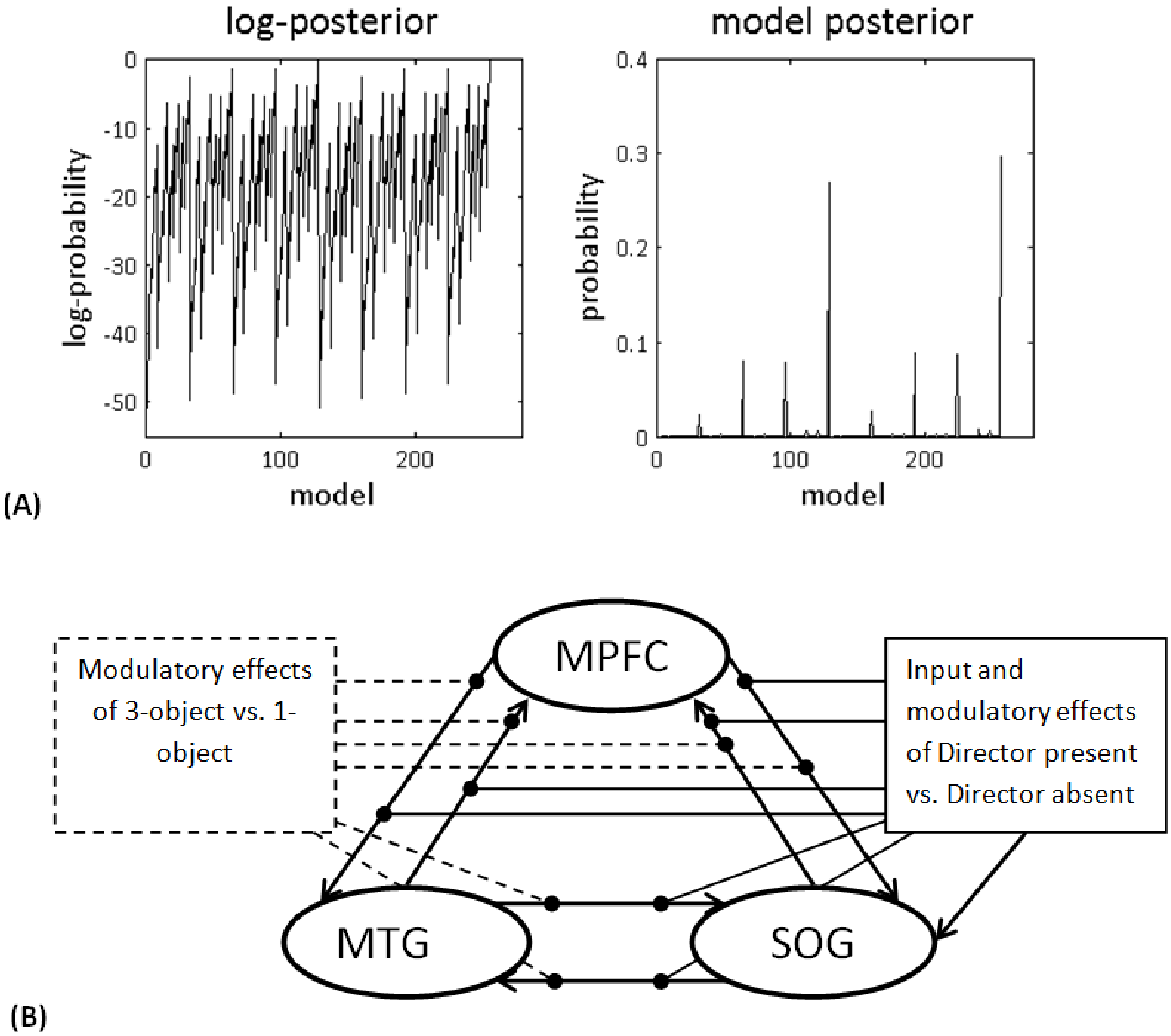


Figure 2: (A) The left graph shows the range of log-posterior probability of all possible models examined. The right graph shows the posterior probability of all models. Model 256 had the highest posterior probability of 0.3. (B) The winning model: The full model was the winning model, with the highest evidence: in this model all connections were modulated by both modulators. The driving input was fixed and went into the most posterior region, the SOG.

Figure 3A (color for the Web)

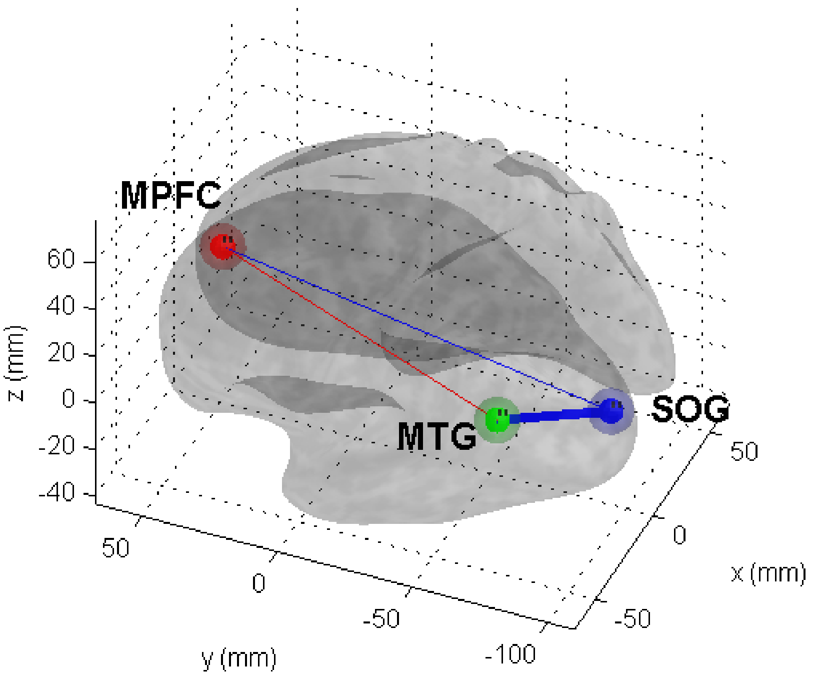


Figure 3: (A) VOIs used in the DCM analyses and illustration of the fixed connectivity between them. Two VOIs were based on the peaks of the Director present > Director absent main effect contrast and were in the superior occipital gyrus (SOG; -12 -94 4; circled in blue) and the left middle temporal gyrus (MTG; -45 -70 13; circled in green). One VOI was based on the PFC peak from Director × Object contrast [(Director present 3-object > 1-object) – (Director absent 3-object > 1-object), inclusively masked by (Director present 3-object > 1 –object)] and was located in the left dorsal medial prefrontal cortex (MPFC; -9 38 34; circled in red). The colour of the line represents the source of the strongest bidirectional fixed connection, while its width represents its absolute strength (see web version for a colour figure of this graph).

Figure 3B (color for the Web)

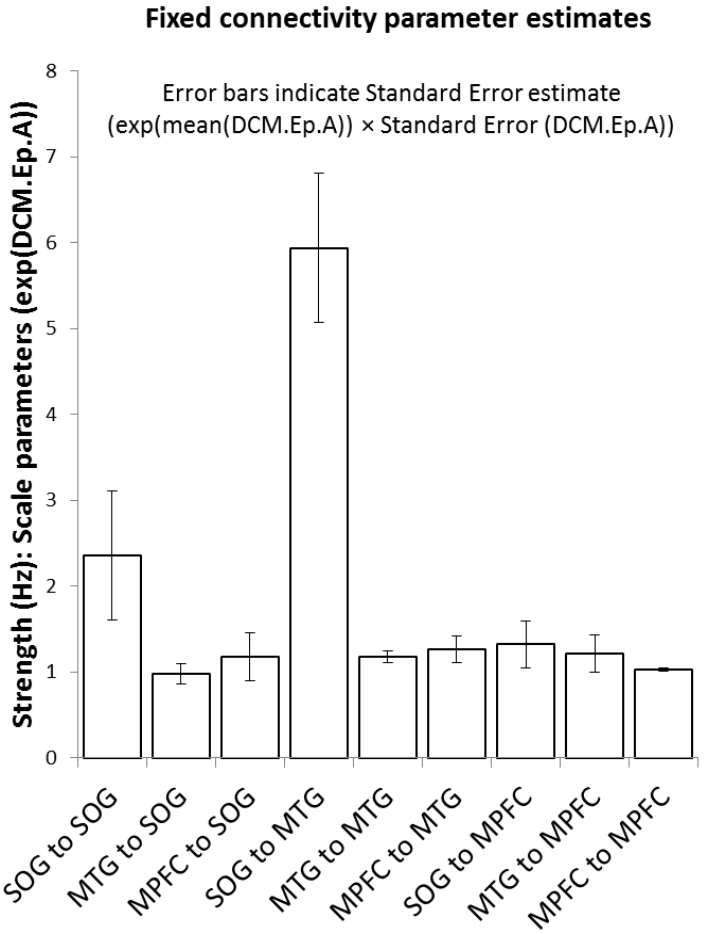
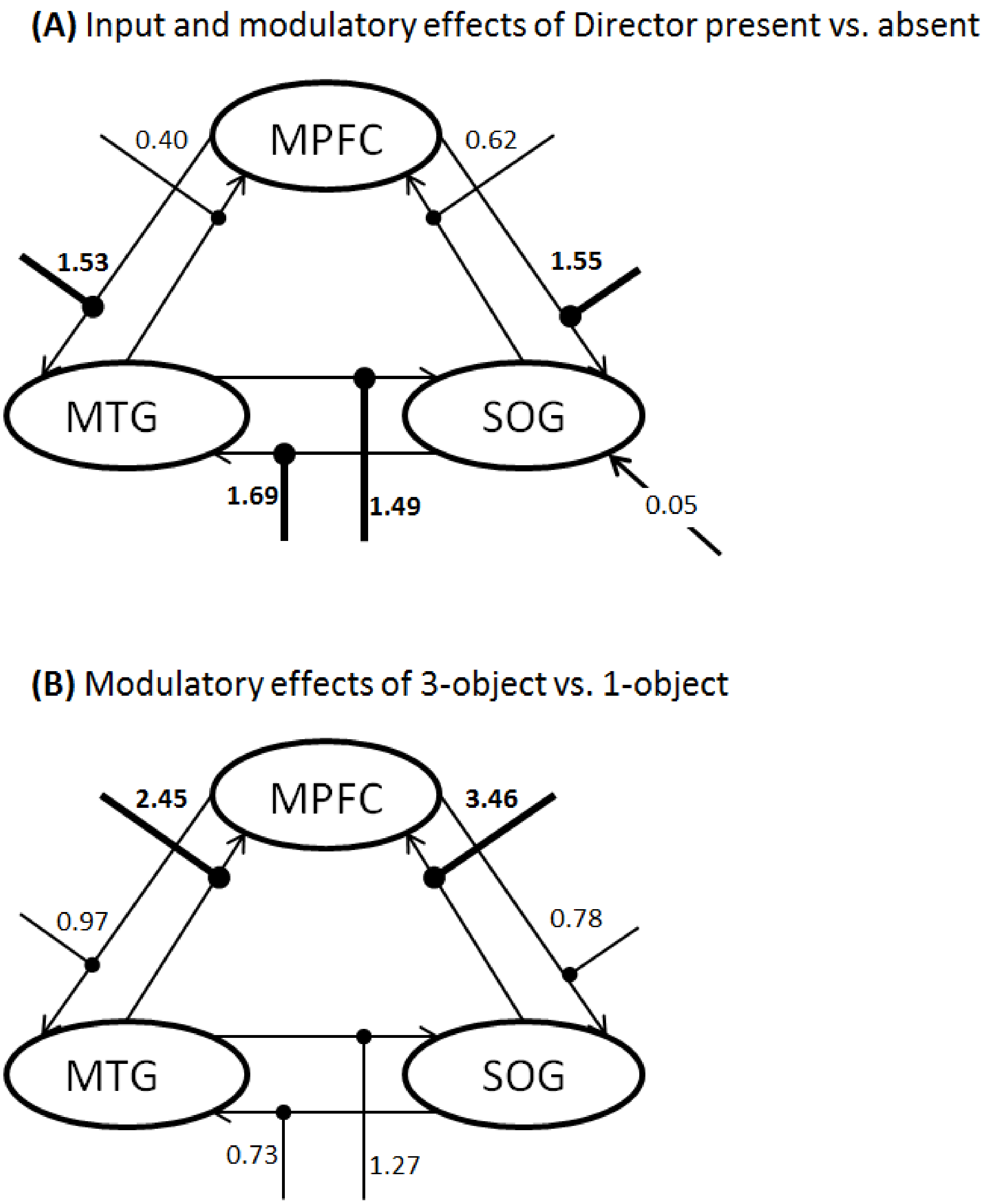


Figure 3: (B) Bayesian Parameter estimates under selected (winning) model of fixed connections strength in absence of task specific modulation. Values indicate scale (exponentiated) parameter estimates for each particular connection.

Figure 4 (color for the Web)



(A) Input and modulatory effects of Director present vs. absent

(B) Modulatory effects of 3-object vs. 1-object

Figure 4: Modulatory connectivity. (A) Modulatory effects of the Director present vs. absent contrast: Backward connections coming from the MPFC (highlighted in bold) and lateral connections between SOG a MTG are subject to stronger modulation. (B) Modulatory effects of the 3-object vs. 1-object contrast: connections leading to the MPFC (highlighted in bold) are

subject to stronger modulation. Values indicate scale (exponentiated) parameter estimates for each particular connection. On (A), the input strength to the SOG is shown. The modulatory parameters estimates can be interpreted as percentage change in connection strength (e.g. the model with the highest evidence has a parameter estimate of 1.55 for the modulation of the MPFC-SOG connection, which corresponds to an increase in MPFC-SOG coupling of 55% during the Director present condition).

Figure 1 (black-and-white for print)

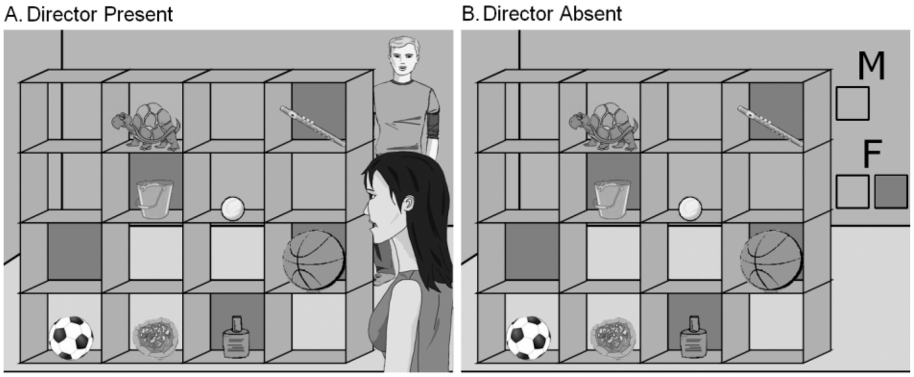


Figure 2 (black-and-white for print)

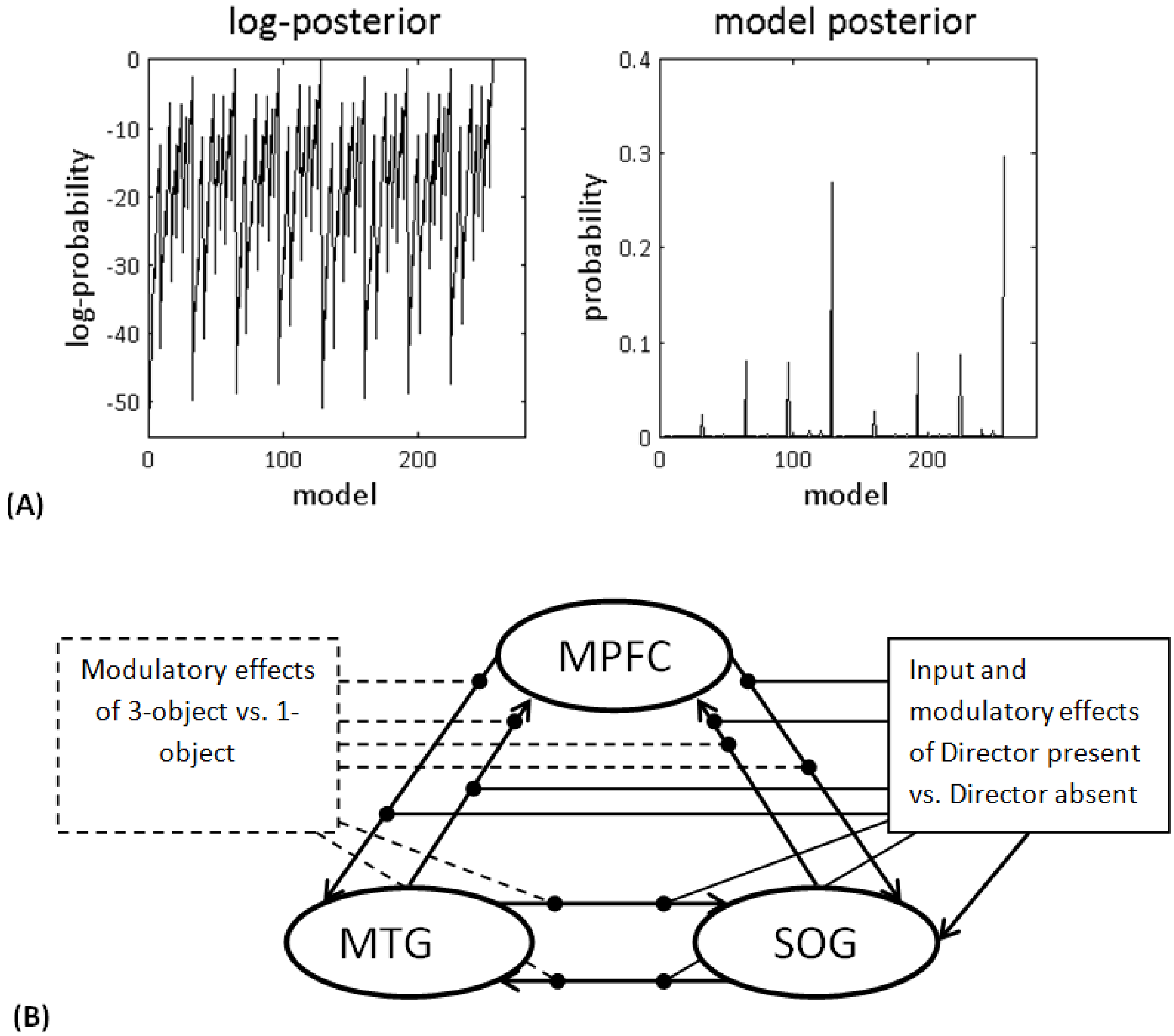
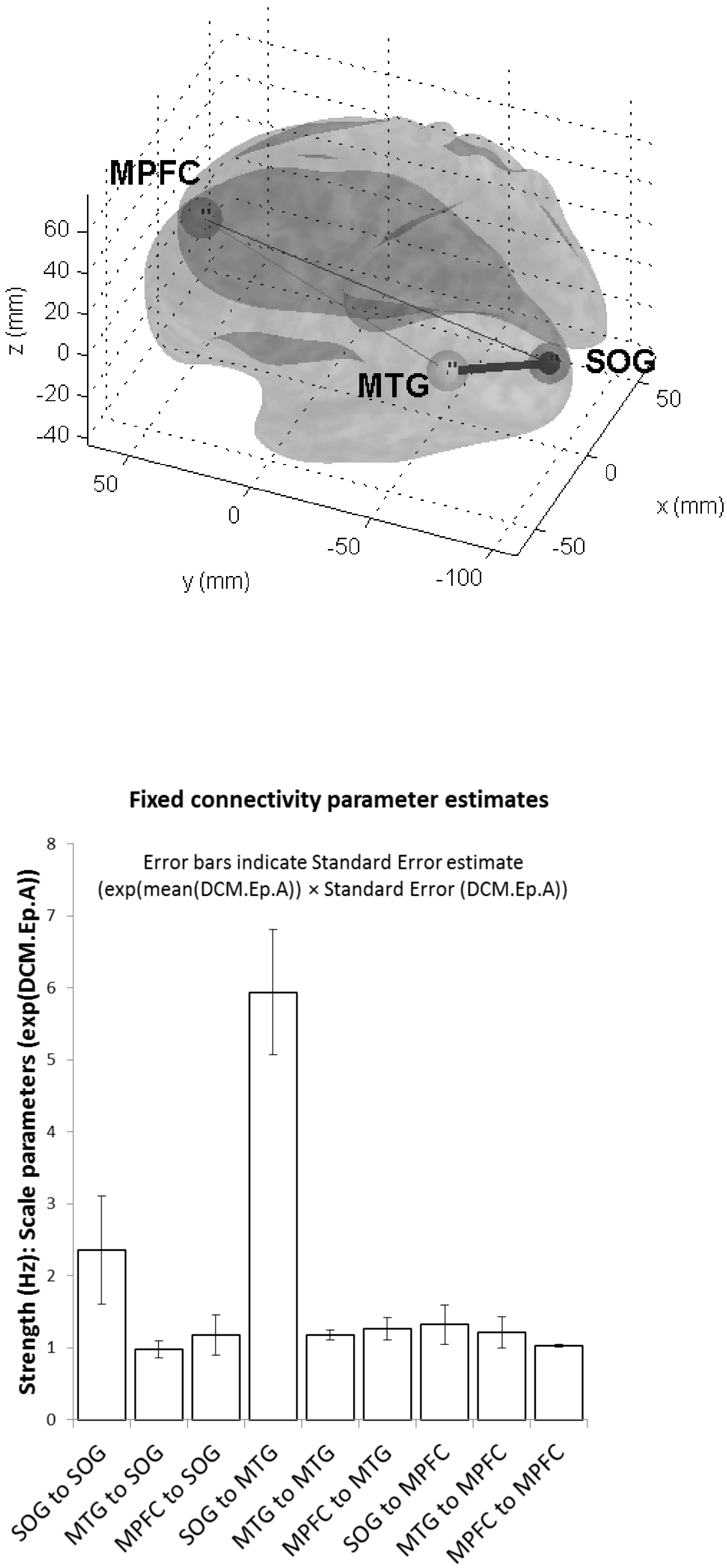


Figure 3 (black-and-white for print)



5. Dynamic Causal Modelling of animacy perception: a Human Connectome study

5.1. Introduction

Humans are highly sensitive to detecting animacy in other agents, noticing entities that act purposefully in the world, a capacity with adaptive benefits for behaviour (Barrett, 2000). The attribution of agency is so automatic and irresistible that humans attribute agency even to simple two-dimensional shapes that appear to move in a self-propelled way (Heider and Simmel, 1944). Cues that trigger the perception of agency include non-Newtonian motion and sudden changes in motion direction and speed (Tremoulet and Feldman, 2000). Animate motion is less predictable than inanimate motion because it is more complex and nonlinear, and because self-propulsion relies on hidden internal (intentional) causes. In contrast, inanimate object motion is generated by relatively invariant forces such as a stone falling due to gravity, or a ball being hit by a snooker cue. Due to the inherent unpredictability, viewing of animate motion compared with inanimate movement should produce more salient or informative sensory prediction error signal (a mismatch between the model's prediction and the sensory input). Here, we test this hypothesis by analysing fMRI data from 132 participants from the Human Connectome Project (HCP, Van Essen et al., 2013) to quantify directed (effective) connectivity during animate versus inanimate motion perception.

fMRI data were collected while participants viewed triangles designed to move in either an animate or inanimate way and Dynamic Causal Modelling (DCM) was used to model the induced responses. DCM estimates the experimental modulation of (intrinsic) self connections or (extrinsic) forward and backwards connections between brain regions that are active during a particular task in a directional manner. We used a novel post-hoc model selection routine (Friston & Penny, 2011; Rosa, Friston, & Penny, 2012) to investigate all possible dynamic causal models, and tested the hypothesis that forward connections, which convey sensory prediction error signals, are selectively more engaged when people view animate movement compared with inanimate physical movement, than top-down backward connections. We quantified the effective connectivity between V5, which is responsive to any type of motion (animate and inanimate) and posterior superior temporal sulcus (pSTS), which is selectively activated when participants view animate motion (Castelli, Happé, Frith, & Frith, 2000).

We investigated whether augmented pSTS responses during the perception of animacy are mediated by a selective increase in sensitivity of forward afferents from motion sensitive V5, or whether they reflect a non-specific increase in pSTS excitability mediated by top-down effects. We expected both an increase in the forward effective connectivity from V5 to pSTS and a decrease in the inhibitory self- or intrinsic-connectivity within pSTS. In addition, we predicted that forward connectivity would be modulated more strongly than the backward connections, due to increased prediction error, attentional selection or gain modulation (Feldman & Friston, 2010).

5.2. Material and methods

Some of the following methods are based on the methods in chapter 4. Where appropriate, we briefly restate them, to allow for easier reading.

5.2.1. Participants

134 healthy adults were initially considered for DCM analysis (89 female, 43 male; all but one participant were between 22-35 years old with a mean age of approximately 30.5 years, see Van

Essen et al. (2012) for why reporting of exact ages would endanger anonymity of participants). Two participants were excluded: one because of missing onset files; another participant did not show activation in the one of the areas during one of the sessions and thus could not be included in the DCM analysis (see *Volume of Interest Extraction section below*). Thus, data were analysed from 132 participants. Participants gave written informed consent (see Van Essen et al., 2013 for more information about ethics) and the data analysis complied with ethical guidelines of the University College London ethics committee.

5.2.2. fMRI data acquisition

See Ugurbil et al. (2013) for a detailed description of the HCP fMRI acquisition protocols. The following abbreviated overview is taken from Barch et al. (2013). Briefly, whole-brain EPI acquisitions were acquired with a 32 channel head coil on a modified 3 T Siemens Skyra with TR = 720 ms, TE = 33.1 ms, flip angle = 52°, BW = 2290 Hz/Px, in-plane FOV = 208 × 180 mm, 72 slices, 2.0 mm isotropic voxels, with a multi-band acceleration factor of 8 (Feinberg et al., 2010; Moeller et al., 2010; as cited in Barch et al., 2013). Two runs of the task were acquired, one with a right-to-left and the other with a left-to-right phase encoding (Barch et al., 2013).

5.2.3. fMRI data preprocessing

We used the “minimally processed” from the Q2 release of the HCP data for this study (Functional Pipeline v2.0; Execution 1). This time series data are preprocessed using tools from FSL and FreeSurfer to implement gradient unwarping, motion correction, fieldmap-based EPI distortion correction, brain-boundary-based registration of EPI to structural T1-weighted scan, non-linear (FNIRT) registration into MNI152 space, and grand-mean intensity normalization (Barch et al., 2013). See Glasser et al. (2013) for a detailed description of fMRI preprocessing of the HCP.

5.2.4. Experimental design

The following abbreviated overview is taken from Barch et al. (2013). A well validated task was used to probe animacy and agency detection, given evidence that such stimuli generate robust task related activation in brain regions associated with social cognition that are reliable across subjects (Castelli et al., 2000, Castelli et al., 2002, Wheatley et al., 2007 and White et al., 2011 as cited in Barch et al., 2013). Participants viewed short video clips (20 s) of objects (squares, circles, triangles) either interacting in some way, or moving mechanically (Barch et al., 2013). The basic visual characteristics in terms of shape, overall speed, and orientation changes were matched between stimulus categories (Castelli et al., 2000). After each video clip, participants rated the video by choosing from three different options, depending on whether the objects contained a social interaction (an interaction that appears as if the shapes are taking each other's feelings and thoughts into account), Not Sure, or No interaction (i.e., there is no obvious interaction between the shapes and the movement appears random). Each of the two task runs comprises 5 video blocks (2 Animate and 3 Inanimate in first run, 3 Animate and 2 Inanimate in the other run) and 5 fixation blocks, which always followed the video blocks (15 s each). Of note, the video clips were shortened to 20 s (the Castelli et al. (Castelli et al., 2000) clips were originally 40 s) by either splitting the videos in two or truncating them. A pilot study by Barch et al. (2013) confirmed that participants rated these shorter videos similarly. **Figure 1** shows stills as an example of Animate motion video.

5.2.5. fMRI Data analysis

fMRI data were further analysed by us, using Statistical Parametric Mapping (SPM12b, www.fil.ion.ucl.ac.uk/spm). The 2 × 2 × 2 mm minimally preprocessed images were spatially smoothed with a 4-mm Gaussian kernel to increase the signal to noise ratio, while retaining sufficient anatomical acuity for extracting visual sensory areas. We did not slice time correct the data, nor did we later specify different acquisition times in the DCM model (Kiebel, Klöppel, Weiskopf, & Friston, 2007), as simulated DCM data has been shown to cope well with slice timing differences up to 1 s (Kiebel et al., 2007) and our TR was 0.72 s. The time series were modelled with boxcar regressors based on two types of task blocks, animate motion and inanimate motion. With these task blocks we created parametric modulators that could mimic conventional *t*-contrasts with single regressors (Büchel, Holmes, Rees, & Friston, 1998) that were orthogonal to each other (the first regressor was All Motion over implicit baseline and the second regressor was Animate – Inanimate motion). This allowed us to use the All motion contrast as a single input to the DCM and the Animate – Inanimate motion contrast as a modulator of effective connectivity. In addition, we included constant session effects. Appropriate stimulus functions were convolved with the canonical hemodynamic response function to form regressors for standard SPM analyses. Together with regressors representing residual movement-related artifacts and their derivatives, these regressors comprised the full (general linear) model (GLM) for each session. A group level ANOVA was performed to identify significant regional effects for the All Motion contrast and a contrast for Animate – Inanimate motion. All analysis scripts are available online (https://github.com/HaukeHillebrandt/SPM_connectome); this ensures the analyses reported below can be replicated and extended with the openly available HCP data (see discussion).

5.2.6. Dynamic causal modelling

DCM estimates the experimental modulation of (intrinsic) self connections or (extrinsic) forward and backwards connections between brain regions that are active during a particular task in a directional manner. This enables one to infer whether experimental manipulations affect top-down, bottom-up influences or both. We used a novel post-hoc model selection routine (Friston & Penny, 2011; Rosa et al., 2012) to investigate all possible dynamic causal models, and tested the hypothesis that forward connections, which convey sensory prediction error signals, are selectively more engaged when people view animate movement compared with inanimate physical movement, than top-down backward connections. Specifically, we quantified the effective connectivity between V5, which is responsive to any type of motion (animate and inanimate) and pSTS, which is selectively activated when participants view animate motion (Castelli et al., 2000).

5.2.7. Specification of dynamic causal models

We created and estimated DCMs (Friston, Harrison, & Penny, 2003) with DCM12 (version 5370) as implemented in SPM12b. The DCMs were based on the VOIs (Volumes of interest) reported above and used the main effect of Animate – Inanimate motion to modulate the connections between regions (see **Figure 3A**). All DCMs were deterministic (as opposed to stochastic for DCMs without experimental input, see (Daunizeau, Stephan, & Friston, 2012), bilinear (as opposed to nonlinear DCMs, where activity between two regions is modulated by a third region, see (Stephan et al., 2008), two-state models (Marreiros, Kiebel, & Friston, 2008), with mean-centred inputs. Our unexponentiated modulatory parameter estimates ranged from -2.7 to 3.9 Hz, similar to one-state DCM parameter estimates reported in other studies (Goulden et al., 2012; Rosa et al., 2012). While the two state-DCMs use exponentiated scale parameters that introduce positivity constraints and are so more plausible to interpret, these values are likely not to be normally distributed and heteroscedastic, because the exponential function is the inverse function of the natural logarithm (which is commonly used to transform data to meet the assumption of a normal distribution, see (Bland & Altman, 1996a, 1996b). Thus, we used the original unexponentiated non-scale parameter estimates for these statistics and plots.

5.2.8. Post-hoc Bayesian model selection

Until recently, it was computationally expensive to estimate a large number models with DCM (Friston, Mattout, Trujillo-Barreto, Ashburner, & Penny, 2007), especially with a large number of participants, as in the current study. A model space with n nodes has $2^{n \times n}$ permutations of connections that can be turned on or off, which can be modulated by different experimental manipulations, leading to a combinatorial explosion (Friston & Penny, 2011). We used a new method to find the model evidence for all possible models by only inverting (estimating) the full model (Friston, Li, Daunizeau, & Stephan, 2011; Friston & Penny, 2011; Rosa et al., 2012) to select the winning model. These post-hoc routines and the conventional variational free energy approach have been shown to yield very similar results (Rosa et al., 2012). First, we used this post hoc model selection procedure (Rosa et al., 2012) to identify the best model out of all possible connection architectures with Bayesian model selection (BMS). Second, we looked at Bayesian parameter averages (BPA) over all possible models showing whether fixed connections existed and whether they were modulated. BPA computes a joint posterior probability density over parameter estimates for a group of participants, by using the posterior from one participant as the prior for the next participant, whose posterior then serves as the prior for the next participant etc. (Kasess et al., 2010; Stephan et al., 2010).The posterior probability is the probability that a model (or family of models) provides the best explanation for the measured data across participants (Penny, Stephan, Mechelli, & Friston, 2004). The probabilities for all analysis were pooled in a fixed effects fashion, because we assumed that the underlying model structure did not vary across the participants. This is similar to family level inferences looking posterior probabilities over parameters (with and without a given free parameter) of whether a fixed connection or a particular modulation exists using family-level inferences (Penny et al., 2010). The post-hoc optimisation also provides parameter estimates for individual participants that can be compared with conventional frequentist statistics (Stephan et al., 2010). Thus we present the simple average parameter estimates for the model with the highest evidence (the winning model) to elucidate the quantitative nature of the connection e.g. how much a connection is modulated (Friston et al., 2003).

5.2.9. Volume of Interest Selection

To identify and summarise regional responses for further dynamic causal modelling we use standard procedures (Hillebrandt, Dumontheil, Blakemore, & Roiser, In press also see online methods). Timeseries from VOIs associated with the above contrasts were summarised using the SPM12b Eigenvariate toolbox: we extracted each participant's principal eigenvariate around the participant-specific local maxima activation nearest to the peak voxel of the group (between subject) GLM analysis (see **Table 1**). The radius of the VOI spheres was 6 mm and the search radius for local maxima from the group analysis was restricted to 20 mm. All voxels contributing to the eigenvariates were significant at $p < 0.05$ uncorrected and adjusted at $p < 0.05$ for the effects of interest (i.e. only for those regressors that were used in the DCMs for input or modulation). We created four separate DCMs (one for each of the two hemisphere and two sessions in order to replicate the results; see **Figure 3B** for a schematic of the model). For each model, the first volume of interest (VOI) was based on maxima in the most active cluster of the All motion contrast (which is Animate and Inanimate motion over the implicit baseline, i.e. unmodelled fixation cross). These maxima were assigned by the SPM anatomy toolbox (Eickhoff et al., 2005) to MT+/V5 (hOC5; right: 44 -64 4; left: -44 -74 4; see **Table 1** for GLM results). V5 was the most active region in our All Motion contrast, and has been shown to be highly sensitive to visual motion (Born & Bradley, 2005). The second VOI was extracted at a local maximum of the most active cluster in each hemisphere based on the results of a conjunction analysis (Price & Friston, 1997). The conjunction was the effect of All Motion & Animate – Inanimate motion to consider areas more active in Animate vs. Inanimate motion, but only in motion sensitive areas (activated by any type of motion). The second VOI (pSTS; sometimes called IPC (PGa and PFm), right: 54 -50 16, left: -56 -52 10) was extracted from the posterior superior temporal sulcus – which was highly active bilaterally: the peaks were local maxima in the most active cluster of each hemisphere with *t*-values above 8. The pSTS has been frequently implicated in animate motion processing in neuroimaging (see discussion). Note that V5 was not significantly more active in this contrast, suggesting that the stimuli were indeed well matched in terms of low-level motion properties. Finally, V5 and pSTS have been shown to have strong (and reciprocal) anatomical connectivity (Lewis & Van Essen, 2000).

5.2.10. Specification of dynamic causal models

Our particular interest was in the effect of animate motion processing on connections among sources in the distributed visual hierarchy. In particular, we wanted to know whether the effect of biological motion processing could be explained by changes – mediated by perceptual set – in intrinsic and extrinsic connections. Furthermore, if these changes were in extrinsic connections, were they in the forward or backward connections? To answer these questions, we used Bayesian model comparison (BMS) of reduced models following inversion of a full model specified as follows: The full model comprised reciprocal connections between the motion sensitive area MT+/V5 and pSTS. The driving input into the model – represented by the DCM.C matrix (Friston et al., 2003) – was the effect of All motion (Animate and Inanimate motion movement, modelled as a single regressor for both types of motion). This driving motion input entered either the posterior, lower region, V5 and modelled extrageniculate input, or the higher cortical node pSTS. Our hypothesis was that V5 would be the first region showing sensitivity to the presence of motion, and this would result in higher parameter estimates for V5 over pSTS as input region. V5 would subsequently influence activity in the pSTS region, but more so in the animate movement condition. These two cortical nodes were reciprocally coupled with extrinsic forward and backward connections, while intrinsic (self or recurrent) connections were treated as inhibitory. The effect of animacy was allowed to modulate all extrinsic and intrinsic connections. This full model was inverted for all subjects and the resulting posterior densities over the connection strengths were used to perform a family wise Bayesian model comparisons using post hoc optimisation. The models considered correspond to all possible combinations of the 10 free coupling parameters – corresponding to 2^10 =1024 reduced models. The 10 parameters comprised 2 fixed intrinsic parameters, 2 fixed extrinsic parameters, 4 parameters controlling the modulation of fixed connectivity and 2 parameters controlling the driving effect of All motion. To examine the connectivity in quantitative terms, we then analysed the posterior distribution over connections under the model with the highest evidence; using both the distribution of estimates over subjects.

In summary, the effect of perceptual set (animate motion) was allowed to change the intrinsic and extrinsic connectivity throughout the hierarchy. We then tested a series of reduced models comparing the evidence for (changes in) intrinsic connectivity, extrinsic connectivity or both. The evidence for these different hypotheses or models was assessed using a variational free energy approximation based upon the post hoc optimisation of reduced versions of the full model. Having identified the model with the greatest evidence, we then characterised the effects of motion and animate motion processing quantitatively, by examining the connection strengths and their bilinear modulation (the model can be replicated with scripts available online – see discussion).

5.3. Results

5.3.1. Behavioural Results

Participants were able to judge whether motion was designed to be animate vs. inanimate as evidenced by the high accuracy levels of their responses (Correct responses: Animate condition: Session 1: $M = 0.90$, $SD = 0.23$, Session 2: $M = 0.96$, $SD = 0.13$; Inanimate condition Session 1: $M = 0.83$, $SD = .21$; Session 2: $M = 0.84$, $SD = .26$).

5.3.2. General Linear Model results

The All motion contrast (against a fixation cross as an implicit baseline) showed increased BOLD signal in many regions (whole-brain voxel-level FWE-corrected threshold of $p < .05$), likely due to very high power. In **Table 1**, we show the most significant activation with a *t*-value of 31 and above. The conjunction analysis of All Motion & Animate – Inanimate motion showed most activity in left and right middle temporal gyri (see **Table 1**).

5.3.3. Dynamic Causal Modelling results

DCMs contained VOIs described above: V5 and pSTS in the right (model 1) and left hemisphere (model 2). The post-hoc analysis (see Methods for details) finds the best model and furnishes the posterior probability of whether individual parameters exist or not. The latter is equivalent to family comparison, which tests whether a family of models without a certain parameter (e.g. a connection between two areas) has a higher probability than the family with this parameter

(Penny et al., 2010). Finally, we compare the strength of connections by examining the winning models parameter estimates.

5.3.4. Bayesian model selection and Bayesian Parameter averaging

We first assessed the model with the best evidence (a metric in which model fit is traded off against model complexity). Comparisons of the evidence for all possible 1024 models showed that the winning (optimal) model with the highest probability had a probability of (almost) 1 (**Figure 2**). The winning model in all four cases (2 (sessions) × 2 (hemispheres)) was always the full model that had all connections and all modulations (**Figure 3A)**. This model has 10 free parameters describing the extrinsic and intrinsic connections and how these connections change with perceptual set. The profile of model (log) evidences over the ensuing 1024 models for one of the four cases is shown in **Figure 2** (other plots were similar), suggesting that the full model had more evidence than any reduced variant. The next most probable model's probability was very low, corresponding to a highly significant difference (Kass & Raftery, 1995). Additionally, this full connectivity was confirmed using Bayesian Parameter averaging (BPA; see Kasess et al., 2010 and supplementary methods), which also showed that all (self) connections and their modulation by animacy were evident with a posterior probability of (almost) 1. Although there was very high evidence for all (self) connections and their modulations, this evidence does not reflect the relative strength of the effects (effect sizes). The fact that we obtained strong evidence for all effects reflects the large sample size. To address specific hypotheses about the locus of animacy effects, we analysed the distribution of parameter estimates under the winning model.

5.3.5. Comparison of connection strength

Since the full model had the highest evidence, we used its parameter estimates to find out which (self) connections were modulated more strongly by animacy. We conducted two separate 2 (hemisphere: right, left) × 2 (session: 1,2) × 2 (connection type in first ANOVA: forward V5-pSTS vs. backward pSTS-V5 connection; connection types in second ANOVA: V5 self-connection, pSTS self-connection) repeated measures ANOVAs for the modulator (Animate – Inanimate motion) on the individual participants' parameter estimate values of the winning model. The results were consistent and clear; showing a large and selective effect of connection type across subjects, hemispheres and sessions, such that the forward connection from V5 to pSTS is more strongly modulated by animacy than the homologue backward connection ($F(1,131) = 667.88$, $p < 0.001$) (see **Figure 4 upper panel**). In the lower panel of **Figure 4,** we show that the intrinsic self-connections of pSTS being significantly lower than V5 ($F(1,131) = 27.47$, $p < 0.001$). Here, we can clearly see that the inhibitory self-connection in pSTS has decreased more than the inhibitory self-connection in V5. In other words, since the inhibitory self-connection is decreased, pSTS activation is augmented by animacy.

5.4. Discussion

In the current large-scale Connectome fMRI study, we used DCM to examine effective connectivity between V5 and pSTS in a task where participants viewed animate versus inanimate motion. We used a novel post-hoc model selection routine (Friston & Penny, 2011; Rosa et al., 2012) to investigate all possible dynamic causal models. Our results suggest that there is reciprocal fixed connectivity between V5 and pSTS in both hemispheres that is independent of the type of motion stimuli and driven by all types of visual motion. The results support our hypothesis that there is modulation of the forward and backward connections between these V5 and pSTS in both hemispheres by whether visual motion is animate or inanimate. Crucially, we found that the modulation of the forward connections was stronger than the modulation of the backward connection and that the inhibitory self-connection of pSTS was decreased during the perception of animacy. Thus, our results show that lower-level motion selective areas influence and are influenced causally by a higher-level area responsive to motion trajectories of animate agents more than by movement of inanimate objects.

The pSTS has been frequently shown to be involved in animate motion processing in neuroimaging studies (for a recent review, see Pavlova, 2012). This region's activity and structure predict task performance on biological motion detection tasks (Herrington, Nymberg, & Schultz, 2011; Gilaie-Dotan, Kanai, Bahrami, Rees, & Saygin, 2013), and transcranial magnetic stimulation (TMS) over pSTS disrupts the perception of biological motion (Grossman, Battelli, & Pascual-Leone, 2005). Moreover, the pSTS has recently is active during attention to agentic movement (Gao, Scholl, & McCarthy, 2012; Lee, Gao, & McCarthy, 2012). For instance, Lee et al. (2012) showed that the pSTS activation (with a peak at: 56 −54 16; remarkably close to our coordinate) was more active when participants attended to chasing (animate) compared to mirror (inanimate) motion in statistically identical random motion.

These findings can be interpreted in the framework of predictive coding as an emerging view of localization of brain function that is based on context and prediction – a view that is also becoming increasingly popular in social neuroscience (Brown & Brüne, 2012; Koster-Hale & Saxe, 2013). For instance, it has recently been shown that when participants predict that they will be shown a face, the fusiform form face area will respond just as much when participants are unexpectedly shown a house as when they are shown a face (also see Clark, 2012; Egner, Monti, & Summerfield, 2010). In predictive coding, forward, bottom-up connections have been hypothesized to propagate prediction error signals about (unexpected) sensory information associated with the stimulus from 'lower' (sensory) brain areas to areas that are 'higher' in the cortical hierarchy (Friston, 2010; Rao & Ballard, 1999). Top-down, backward connections then send predictions based on an internal generative model about the stimulus to lower sensory areas to minimize sensory prediction error (see Friston, 2010; Clark, 2012). Furthermore, the selective modulation of prediction errors provides a mechanism for selectively attending to particular prediction errors that inform high-level representations (Feldman & Friston, 2010). In Lee et al's (2012) and our studies, the lower level properties (such as direction changes) of the motion are well matched, but participants' prior expectation to attend more to the animate motion is associated with more activity in the pSTS, potentially because participants' internal model cannot predict animate motion as well. This is also supported by a study showing neural adaptation to repeated exposure of the same animate motion trajectories (Ramsey & Hamilton, 2010) and behavioural evidence showing that more animate triangles invite longer fixation times (Klein, Zwickel, Prinz, & Frith, 2009). Thus, animacy leads to more sensory processing and more rapid updating of initial predictions (Friston, Adams, Perrinet, & Breakspear, 2012). Moreover, the unpredictability of animate motion might give rise to the phenomenological states that accompany social perception and higher-level cognition like attributing agency and free will and thinking about the mental states of an agent.

5.4.1. Future research facilitated by open science

Recently, fMRI studies have been criticized for low statistical power due to relatively small sample sizes (Button et al., 2013; but see Friston, 2012). A recent study reported that fMRI studies with small sample sizes discover as many foci as larger studies, even though more foci should be activated as sample size and thus statistical power increases (David et al., 2013). This is suggestive of a strong reporting bias in the fMRI literature, leading David and colleagues to call for the generation of standardized large-scale evidence in the field (David et al., 2013). Here we analysed data from 132 participants to achieve high statistical power.

Open sharing of task-based fMRI data is becoming increasingly popular and has a wide array of advantages (Kandel, Markram, Matthews, Yuste, & Koch, 2013; Poldrack et al., 2013). The HCP data are publically accessible our analysis scripts used in the current study are freely available online (https://github.com/HaukeHillebrandt/SPM_connectome), which will facilitate replications and extensions of the present findings (Carp, 2013). As more data become available, the analyses can be extended to more subjects. Moreover, as the HCP provides data from other imaging modalities (Van Essen et al., 2012), one can incorporate these into existing models. For instance, one might improve DCMs with tractography-based (Stephan, Tittgemeyer, Knösche, Moran, & Friston, 2009) or electrophysiological priors (Nguyen, Breakspear, & Cunnington, in press). It would be interesting to see whether future DCM extensions can improve the model fit: for instance, nonlinear DCMs where one region can influence connections between other regions, can sometimes outperform bilinear DCMs (Stephan et al., 2008), and future studies could compete in a challenge to find a plausible model that best explains the same data (as it is sometimes done in the machine learning).

5.4.2. Conclusion

DCM provides evidence from data from a very large number of participants for set-dependent changes in the sensitivity of the pSTS to both forward afferents from motion sensitive areas like V5 and recurrent connections within the temporal region. These are likely to be mediated by top-down effects that establish the perceptual set that is engaged during the perception of animate motion. Furthermore, these results speak to the reproducibility and consistency of effective connectivity estimates in a large number of participants and demonstrate the increase in statistical efficiency afforded by large cohorts. Our results show that while both forward and backward connections from V5 to pSTS are more modulated while participants view animate vs. inanimate movement, the forward connection from V5 to pSTS were more modulated than the backward homologue. This suggests that the biological complexity of modelling and predicting movement of other agents leads to higher sensory prediction error.

5.4.3. Acknowledgements

Data for this chapter were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

5.5. Tables

Table 1: GLM results

A) All motion over implicit baseline contrast

The analyses were performed with FWE correction at $p < 0.05$, whole brain level and cluster size of more than 5 voxels. Listed are only those peaks that have a *t*-value of 31 or above. Coordinates used for VOI extraction are in BOLD.

Maximum 01	T = 33.18	MNI:	32	-70	28		Right Middle Occipital Gyrus
Maximum 02	T = 32.41	MNI:	-42	-68	2		Left Middle Occipital Gyrus
	Probability for	hOC5 (V5)			20	%	
Maximum 03	T = 31.97	MNI:	-26	-74	26		Left Middle Occipital Gyrus
Maximum 04	**T = 31.78**	**MNI:**	**44**	**-64**	**4**		**Right Middle Temporal Gyrus**
-> Assigned to right hOC5 (V5). Probability for hOC5 (V5): 40%							
Maximum 05	**T = 31.34**	**MNI:**	**-44**	**-74**	**4**		**Left Middle Occipital Gyrus**
-> Assigned to left hOC5 (V5). Probability for hOC5 (V5): 50%							

B) Conjunction analysis of All motion & Animate – Inanimate motion

The analyses were performed with FWE correction at $p < 0.05$, whole brain level and cluster size of more than 5 voxels. Listed are only those peaks that have a *t*-value higher than the pSTS that were extracted in both clusters. Coordinates used for VOI extraction are in BOLD.

Cluster 1 (2405 voxel): Left hemisphere

Maximum 01	T = 16.71	MNI:	-20	-76	-36	Left Cerebellum -> Assigned to left Lobule VIIa Crus II (Hem)

	Probability for	Lobule VIIa Crus II (Hem)	74%			
	Probability for	Lobule VIIa Crus I (Hem)	22%			
Maximum 02	T = 13.51	MNI:	-20	-72	-28	Left Cerebellum -> Assigned to left Lobule VIIa Crus I (Hem)
	Probability for	Lobule VIIa Crus I (Hem)	50 %			
	Probability for	Lobule VI (Hem)	50 %			
Maximum 03	T = 11.75	MNI:	-42	-50	-14	Left Fusiform Gyrus
Maximum 04	T = 10.45	MNI:	-46	-56	-16	Left Fusiform Gyrus
Maximum 05	T = 10.32	MNI:	-30	-46	-6	Left Lingual Gyrus
Maximum 06	T = 10.15	MNI:	-32	-46	-22	Left Fusiform Gyrus
	Probability for	Lobule VI (Hem)	2 %			
Maximum 07	T = 9.91	MNI:	-34	-54	-18	Left Fusiform Gyrus
Maximum 08	T = 9.34	MNI:	-32	-42	-10	Left ParaHippocampal Gyrus
Maximum 09	T = 9.32	MNI:	-52	-26	-2	Left Middle Temporal Gyrus
Maximum 10	T = 9.16	MNI:	-52	-36	2	Left Middle Temporal Gyrus
Maximum 11	T = 8.99	MNI:	-30	-40	-22	Left Fusiform Gyrus
Maximum 12	T = 8.97	MNI:	-24	-48	-10	Left Lingual Gyrus
Maximum 13	T = 8.89	MNI:	-52	-52	22	Left Middle Temporal Gyrus
Maximum						
Maximum 14	T = 8.62	MNI:	-32	-64	-24	Left Cerebellum
Maximum 15	**T = 7.82**	**MNI:**	**-56**	**-52**	**10**	**Left Middle Temporal Gyrus (pSTS)**

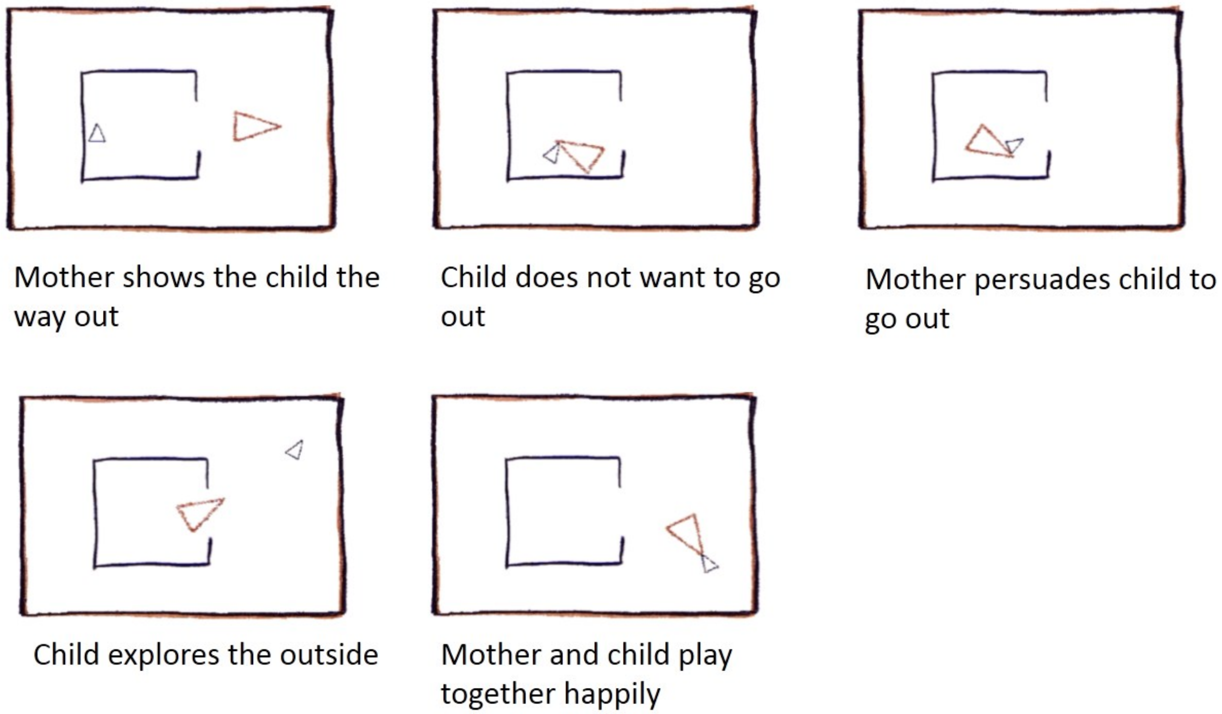
Cluster 2 (1118 voxel): Right hemisphere

Maximum 01	T = 14.44	MNI:	48	-24	-6	N/A
Maximum 02	T = 13.10	MNI:	48	-36	4	Right Middle Temporal Gyrus
Maximum 03	T = 10.81	MNI:	52	4	-18	Right Medial Temporal Pole

Maximum 04	T = 10.53	MNI:	58	-40	6	Right Middle Temporal Gyrus
Probability for		IPC (PGa)		10%		
Maximum 05	T = 9.73	MNI:	56	-32	-2	Right Middle Temporal Gyrus
Maximum 06	T = 9.00	MNI:	50	-4	-16	Right Middle Temporal Gyrus
Maximum 07	**T = 8.72**	**MNI:**	**54**	**-50**	**16**	**Right Middle Temporal Gyrus**
Probability for		**IPC (PGa)**			**30%**	
Probability for		**IPC (PFm)**			**20%**	

5.6. Figures

Figure 1 (color for the Web)



Mother shows the child the way out

Child does not want to go out

Mother persuades child to go out

Child explores the outside

Mother and child play together happily

Figure 1: Example of "Theory of Mind" animation: The Big Triangle coaxing the reluctant Little Triangle to come out of an enclosure (participants do not see captions) (stimuli and description adapted from (Happe, 2001).

Figure 2

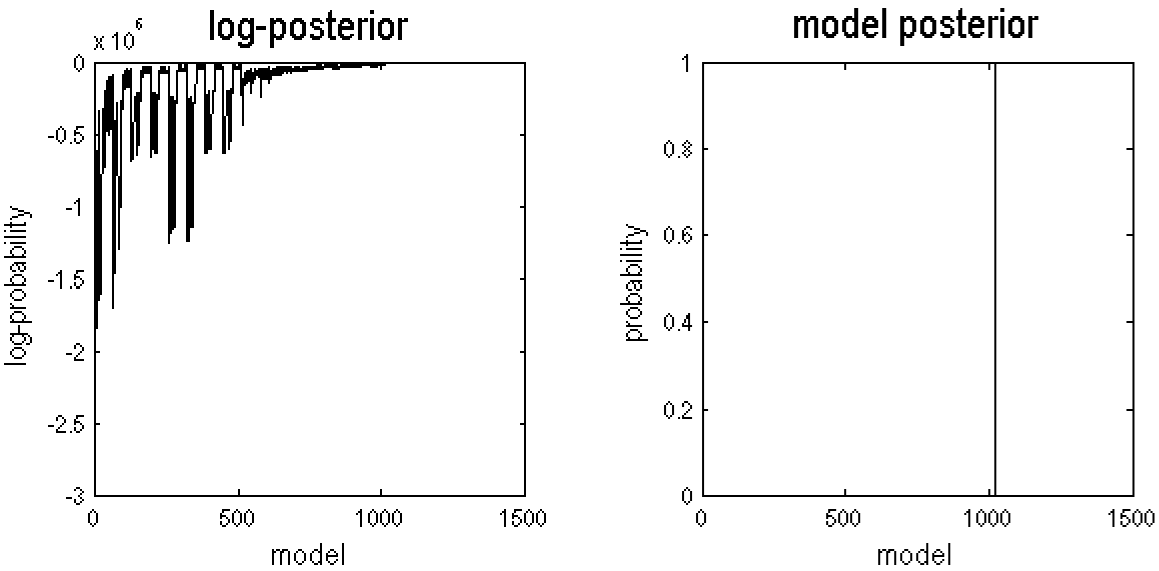


Figure 2: The left graph shows the range of log-posterior probability of all possible models examined. The right graph shows the posterior probability of all models. Model 1024 had the highest posterior probability of (almost) 1. This graph shows data for the first session and the right hemisphere, results for other sessions and hemispheres was similar.

Figure 3A

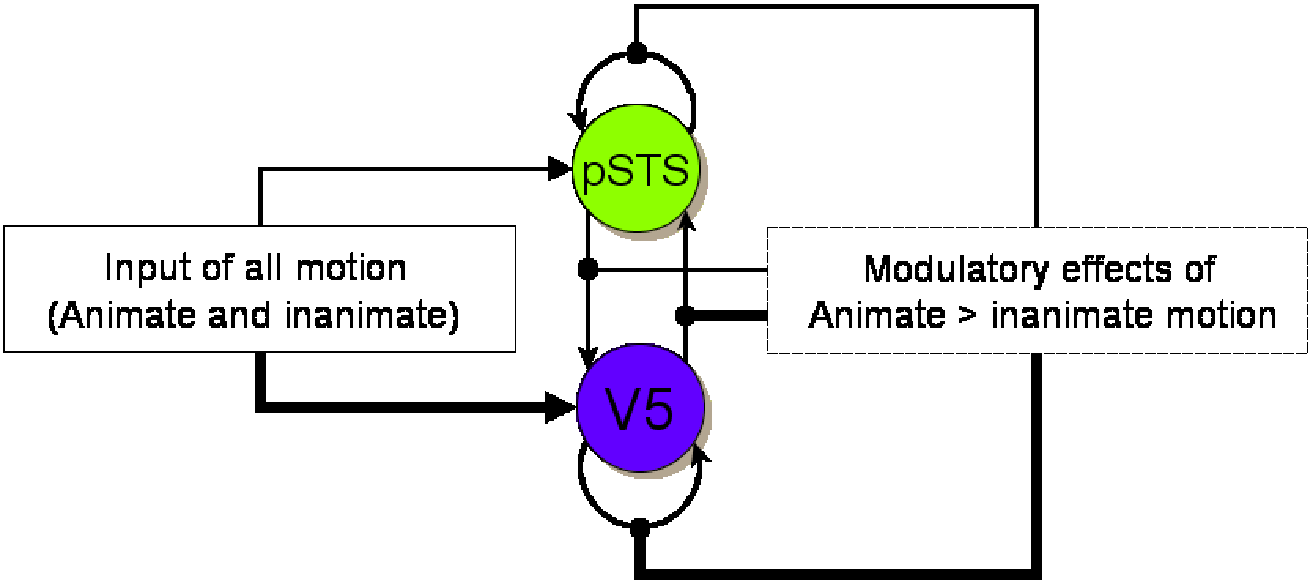


Figure 3: (A) The winning model: The full model was the winning model, with the highest evidence: in this model all connections were modulated by the Animate – Inanimate motion modulator. The driving input was All motion and went into V5 and pSTS. Wider lines represent stronger modulation or input relative to its comparison: V5 received more input than pSTS and the Animate – Inanimate motion contrast modulated the forward connection from V5 to pSTS more strongly than the backward connection and the inhibitory self-connection of pSTS less strongly than the self-connection of V5.

Figure 3B

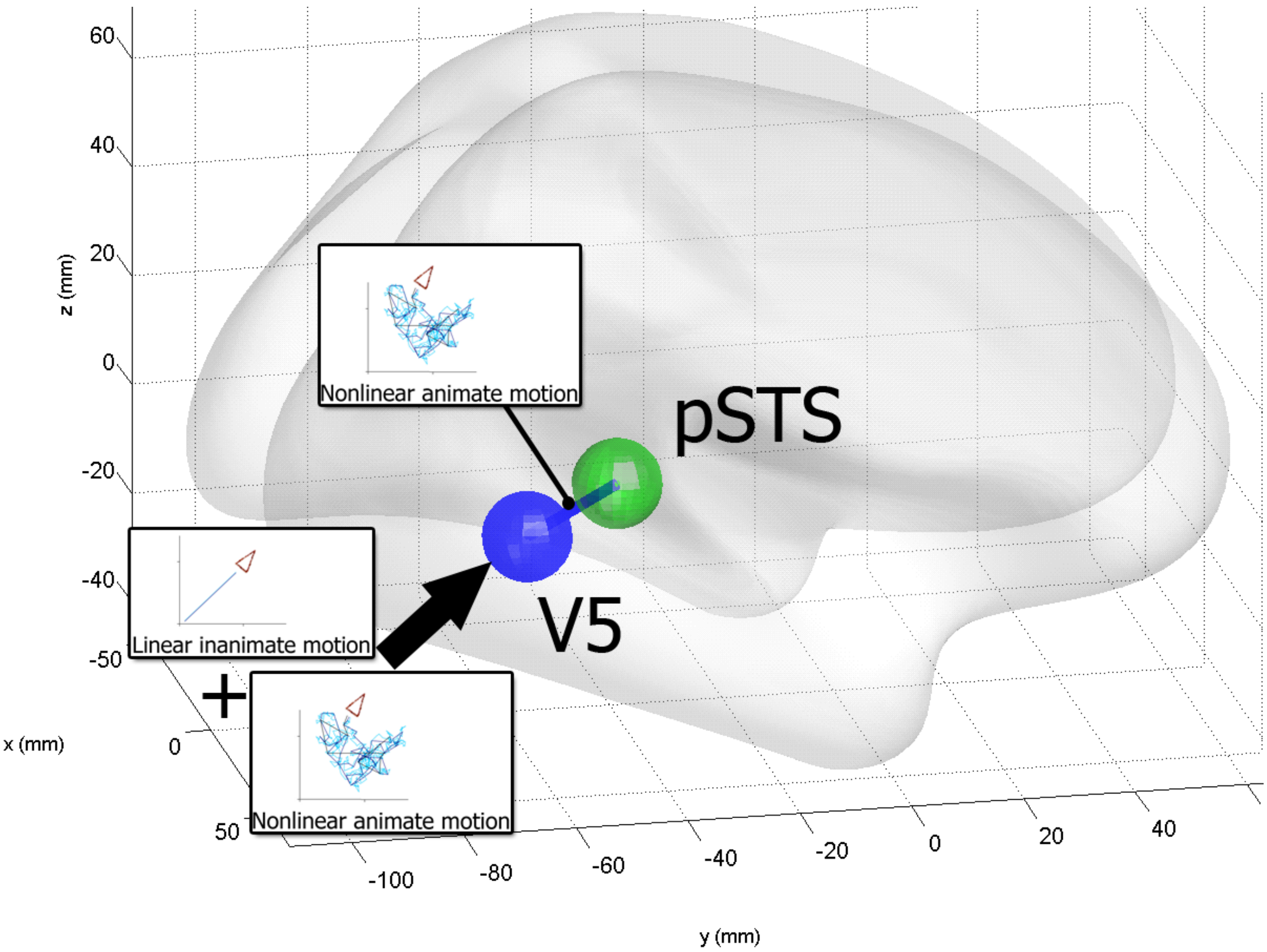


Figure 3: (B) VOIs used in the DCM analyses and illustration of the modulatory connectivity between them. The first VOI was based on the peaks of the All Motion contrast was in the MT+/V5 (44 -64 4; circled in blue). The other VOI was active on the peaks of the conjunction of the All motion contrast and the Animate – Inanimate motion contrast [All Motion & Animate – Inanimate motion] and was located in posterior superior temporal sulcus (pSTS; 54 -50 16, circled in green). The colour of the line represents the source of the strongest bidirectional modulatory connection.

Figure 4

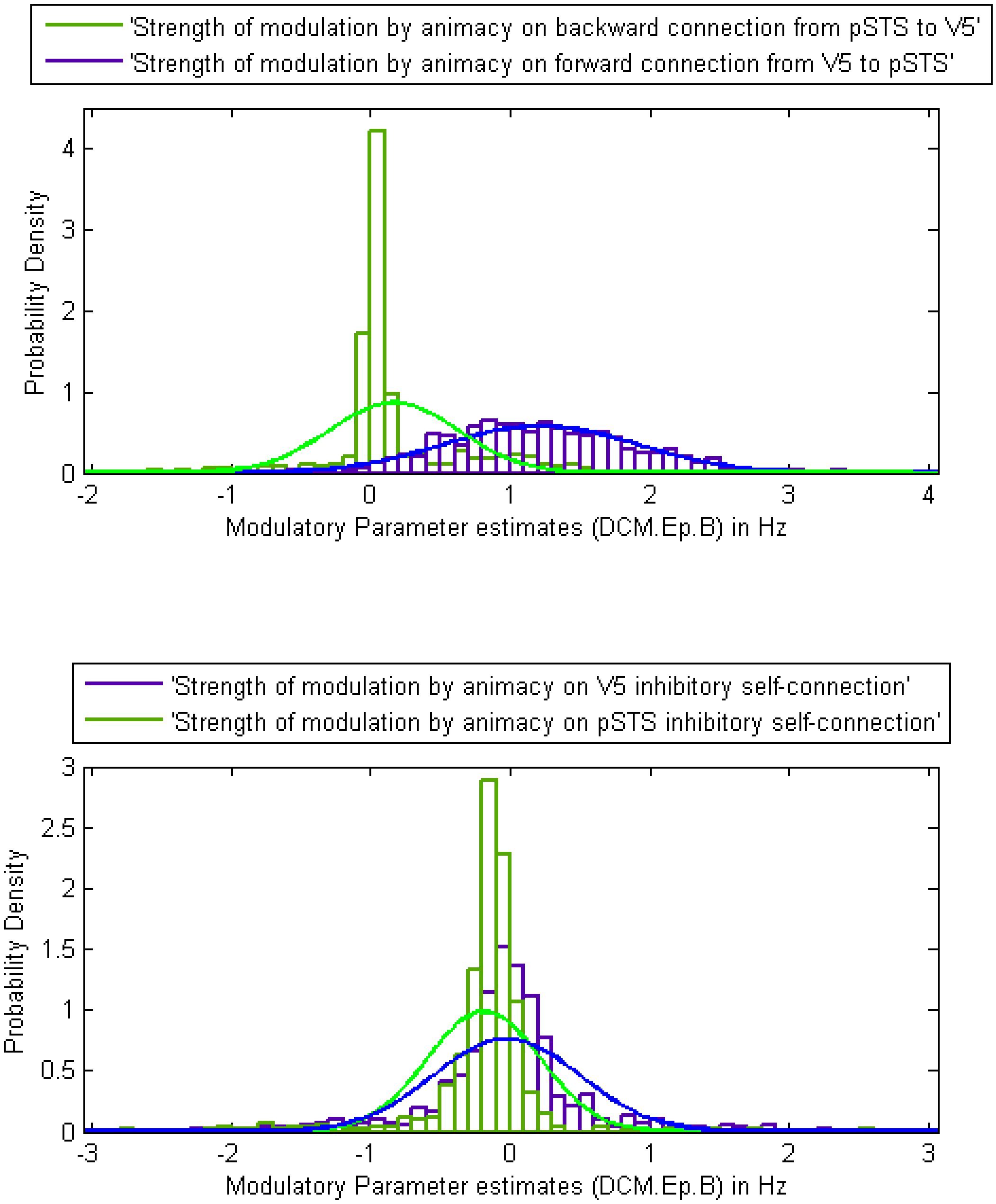


Figure 4: Probability densities functions of parameter estimates for individual participants showed how strongly (self-)connections were modulated by animacy across subjects,

hemispheres and sessions. **Upper Panel:** The forward connection from V5 to pSTS is more strongly modulated by animacy than the homologue backward connection. **Lower Panel:** The intrinsic self-connections of pSTS is significantly lower than V5 and one can clearly see that the inhibitory self-connection in pSTS has decreased towards zero consistently more than the inhibitory self-connection in V5. Since inhibitory self-connection is decreased by animacy, this causes the pSTS activation during the Animate > Inanimate contrast.

6. General Discussion

6.1. Summary and interpretation in predictive coding framework, limitations and future directions

The studies in this thesis attempted to address questions about the (neural) mechanisms of social cognition and social behaviour. The discussion of this thesis will summarize the results and interpret them in the hierarchical predictive coding framework as well as discuss limitations and extensions (future research) of the research project.

6.2. Chapter 2: Experimentally induced social inclusion influences behaviour on trust games

In chapter 2, we experimentally manipulated whether people were socially included or excluded and then measured how much they would trust either complete strangers or the people who included or excluded them. The Inclusion/Exclusion manipulation interacted with Group such that participants in the Reputation-Group (i.e. the group in which participants played with the same players in two games, so that they could base their decisions on a reputation in the second game) trusted individuals who included them more than those who excluded them, whereas inclusion/exclusion made no difference to trust in the No-Reputation-Group (i.e. the group in which participants played the two games with different players, that had no reputation). As had previously been speculated, our findings suggest that exclusion does not increase gullibility. This was evidenced by participants not trusting people who had previously excluded them, to reestablish positive relations. In contrast, reputation is transferred from a social to an economic setting so that social inclusion increases trust.

In a predictive coding framework, the internal model of the participants makes no specific predictions before the social inclusion or exclusion encounter, because on average they have no prior expectations and no model of the other participants that they could use to predict their behaviour (as all the photos of the other players were randomized). This is not to say that they have no predictions of how to approach such cooperative behaviour with strangers. For instance, recently it has been shown that by default people intuitively cooperate with others and this might be advantageous in daily life (Rand, Greene, & Nowak, 2012). Moreover, people routinely display ‘tit for tat’ or ‘win stay, lose shift’ strategies (Nowak & Sigmund, 1993) and these might be the default behaviour for cooperative situations that have been selected for. However, despite such default behavioural strategies, participants take into account whether the strangers exclude or include them socially and then update their (generative) model based on this experience. Upon encountering people who include or exclude them, they update their model of the other person as evidenced by their adapting their trusting behaviour to the context. Interestingly, there was less trust after exclusion than after inclusion in the Reputation group and the trust after exclusion was not different from the level of trust that strangers were met with (in the No Reputation group) - a level of trust that was very similar to was reported in a recent meta-analysis of 84 trust games (Johnson & Mislin, 2008). This suggests that the default model of others is one that where people are seen as not very trust worthy and the default behavioural approach is such that people are cautious to be exploited by people they do not know well. In this light, inclusion by relative strangers causes a prediction error (is unexpected), which leads to a change in behaviour (more trust).

6.2.1. Future Directions and Limitations

The paradigm used in chapter 2 might be seen as an example that paves the way for even more realistic social interactions. For instance, one could use paradigms that are based on more realistic video games (based on commercial video games), where participants have to interact with other players (who could be confederates or bots) and manipulate and measure a manifold of parameters as independent and dependent variables, not limited to the effects of social inclusion and exclusion and trust. In this regard, the research that I presented in chapter 2 is also limited as it might not have complete ecological validity and people might actually behave differently towards people who exclude them in real life. In addition, we have not looked at the neural or other physiological correlates of social exclusion or trust. Research on this is the topic of ongoing investigation: the neural and physiological correlates of trust (for a recent review see Tzieropoulos, 2013) and social exclusion (for a review see Eisenberger, 2012) are being increasingly studied.

6.3. Chapter 3: Developmental Differences in the Control of Action Selection by Social Information

Our everyday actions are often performed in the context of a social interaction. In adults, selecting an action on the basis of either social or symbolic cues was associated with activations in the frontoparietal cognitive control network, while the presence and use of social vs. symbolic cues was in addition associated with activations in the temporal and medial prefrontal cortex (mPFC) social brain network. Here I investigated developmental changes in these two networks. Fourteen adults (age 21-30) and 14 adolescents (11-16) performed an adapted version of the Director Task described above. They followed instructions to move objects in a set of shelves.

Interpretation of the instructions was conditional on the point of view of a visible “director” or the meaning of a symbolic cue (Director Present vs. Director Absent), and the number of potential referent objects in the shelves (3-object vs. 1-object). This study attempts, therefore, to show developmental differences in domain-general and domain-specific PFC activations associated with action selection in a social interaction context. The 3-object trials elicited increased frontoparietal and temporal activations, with greater left lateral PFC and parietal activations in adults than adolescents. Social vs. symbolic information led to activations in superior dorsal MPFC, precuneus, and along the superior/middle temporal sulci. Both dorsal MPFC and left temporal clusters exhibited a Director × Object interaction, with greater activation when participants needed to consider the directors’ viewpoints. This effect differed with age in dorsal MPFC. Adolescents showed greater activation whenever social information was present, while adults showed greater activation only when the directors’ viewpoints were relevant to task performance.

From a predictive coding perspective, it might be argued that adolescents have a less complete model of the social world and thus more prediction error is carried forward, which causes increased activity in higher-level areas like the mPFC. Future work on the social brain in adolescence is currently underway with many different research avenues exploring the neural, physiological and genetic underpinnings of social behaviour during adolescence (Blakemore & Mills, 2014).

6.3.1. Limitations: Group Studies

Functional magnetic resonance imaging (fMRI) measures blood level oxygenation dependent (BOLD) signal and does not measure neuronal activity directly. Thus, changes in BOLD signals may not reflect altered information processing, if there are changes in neurovascular coupling for example. In this case, any comparison between groups (such as different age groups) that have systematic differences in blood oxygenation might influence the results and might not reflect differences in neural activity (Harris, Reynell, & Attwell, 2011). For instance, differences in somatic states such as heart rate can influence the BOLD signal (Chang, Cunningham, & Glover, 2009). Differences in head movements might also play a part and in fcMRI (functional connectivity functional magnetic resonance imaging) motion-related artefacts are thought to affect functional connectivity time course data in rs-fcMRI (resting state functional connectivity functional magnetic resonance imaging) studies (Power, Barnes, Snyder, Schlaggar, & Petersen, 2012). This is potentially not as problematic for group comparisons that use tasked-based fMRI with multifactorial designs and test very constrained interactions. In other words, differences between two different groups in resting state fMRI measures might be due to differences in breathing, head motion that is not controlled for (one does not know about interscan movement), head sizes that cannot be normalized completely, etc. In task-based fMRI, the design is more constrained and some of those confounds become less plausible (although they cannot be completely ruled out). For instance, group differences in a main effect of seeing an angry face vs. a neutral face could be due neurovascular confounds created by more gasping when one group when seeing an angry face. However, as the number of factors increases as in our study, neurovascular confounds (for instance, due to breathing) become more unlikely as one group would have to breathe differently in one particular cell of an interaction, but not others.

6.4. Chapter 4: Dynamic causal modelling of effective connectivity during perspective taking in a communicative task

In chapter 4, we used DCM to explore effective connectivity between SOG, the MTG, and the MPFC, in the Director task from the previous chapter that required participants to take into account another person's perspective in order to guide action selection. We used post-hoc model selection routine to look at all possible dynamic causal models. Family-level inference provided strong evidence for reciprocal fixed connectivity between all three areas and modulations by both the Director and the Object factors, which respectively manipulate the nature of the stimuli (social vs. non-social) and the need for top-down influences on action selection. Crucially, the results show that the social demands modulate the backward connections from the mPFC more strongly than the forward connections from the superior occipital gyrus (SOG) and the medial temporal gyrus (MTG) to the mPFC. This was also the case for the backward connection from the MTG to the SOG. Conversely, the executive task demands modulated the forward connections of the SOG and the MTG to the mPFC more strongly than the backward connections. In predictive coding, the top-down backward connections pass predictions based on an internal (generative) model about the stimulus to lower sensory areas to minimize sensory prediction error (by selectively sampling the stimulus array) and to induce behavioural responses (Clark, 2012; Friston, 2010). In our 3-object factor, the stimulus array has to be sampled more extensively than in the 1-object condition like in traditional top-down visual search paradigms (see for instance, Buschman and Miller, 2007) to find the correct target ('which ball is referred to?) amidst distractors, and so 'surprising' unpredicted sensory information about distractors is passed forward in the cortical hierarchy as prediction error, unlike in bottom-up searches (where a ball 'pops out' among trucks). Lastly, the main effect of director not only modulates particularly strongly the backward connections (potentially by passing down predictions about the director), but also modulates what is usually characterized as a forward connection from

SOG to MTG. Because the task requires multisensory integration of visual stimuli with auditory instructions, the connection from SOG to MTG might be a backward connection where the SOG predicts auditory responses in the MTG in a top-down fashion.

6.4.1. Future directions

Humans are a very cooperative species. In order to cooperate, we have to communicate with each other. The transmission and reception of these communication signals must of course have a neural basis. In the Director task that we have used, we have only modelled the behaviours and neural responses of one of the interaction partners, while standardizing the responses of the director. Using hyperscanning (Montague et al., 2002), a term that is used for scanning the brains of two or more people who are interacting at the same time and looking for relationships in these two people's neural measures, one could create dynamic causal models of social interactions modelling both interactants. For this, one could have two people communicate, while hyperscanning both people. Using DCM, one would be the 'experimental input' to the other person's brain responses, whose brain responses in turn form the input for the first person's brain responses. These models might be complicated and not feasible for technical reasons at present, but eventually it might be interesting to create causal models of social interactions that go above and beyond the search for statistical dependencies in form of correlations between interacting brains. In a hyperscanning DCM paradigm with two people, for instance, one could more completely model the social circuit diagram of smiling at each other. The design might be as follows. The first person might be smiling at the person she is interacting with, while both people are scanned. This smiling is triggered by neural activity in motor cortices of person 1. The first person's motor cortex responsible for a smile might be directly connected to the face recognition area of another person and smiling might modulate the effective connectivity between person 1's motor cortex and person 2's fusiform face area. Then, Person 2's face recognition area is linked to the reward system, the mirror neuron system and motor systems, which by mirroring the smile will activate person 1's face recognition area again, ad infinitum. Modelling these feedback loops might be interesting.

Finally, we have not examined the effective connectivity in the sample of adolescent participants from chapter 3 and compared their connectivity parameter estimates to those of our adult participants. Future research could investigate this question. It would be of particular interest whether connectivity parameter estimates are predictive of age over and above traditional mass univariate and multivariate measures and whether they are able to predict behavioural performance.

6.4.2. Lack of quantitative specification of functionally defined areas (Chapter 3 & 4)

A further problem that is common in (social) cognitive neuroscience and which also affects the first two fMRI studies presented here, is the lack of a standardized (quantified) procedure for assigning anatomical labels to areas (especially for cortical areas). Anatomical labelling of functional activations by humans has the drawback that different researchers disagree about the name of the area that is activated. This leads to some ambiguity in how activations in one study relate to activations in other studies. For instance, different researchers call areas with a wide range of coordinates medial prefrontal cortex (Van Overwalle, 2009) or areas are sometimes specified with many different prefixes, such as dorsolateral medial prefrontal cortex (see for instance, Molina et al., 2011). The resulting flexibility in data analysis might lead to increases in false positives (Simmons, Nelson, & Simonsohn, 2011). In the experiment described in chapter 5, we addressed this problem partially, by using the SPM anatomy toolbox which uses the MNI coordinates and probabilistic cytoarchitectonic maps to label activations (Eickhoff et al., 2005).

6.5. Chapter 5: Dynamic Causal Modelling of animacy perception: a Human Connectome study

Biological agents are the most complex systems humans encounter in their natural environment and their behaviour is highly nonlinear. In uncontrolled, non-experimental settings, they are much more complex than most other non-biological systems (e.g. a tree moving in the wind), and the laws of physics. Primates on the other hand have brains with billions of neurons (Azevedo et al., 2009) that are dynamically interacting. Thus modelling this system (interpreting others' mental states) completely is not only complicated, but also limited by the physics of our brain: modelling another brain exhaustively is impossible because one's model of another person must include the model another person has of one, ad infinitum (Friston, Thornton, & Clark, 2012). However, it is critical to model others' mental states correctly in order to predict their behaviour adequately (if imperfectly) for adaptive success, i.e. for instance, to model the behaviour of predators and prey, and also to model conspecifics' mental states in order to be able to cooperate with them (Barrett, 2000). Consider a human observing a primitive organism performing some taxis behaviour. For instance, Caenorhabditis elegans (roundworm) performs chemotaxis behaviour (moves towards certain chemicals) and one can manipulate this behaviour by increasing the concentration of some chemical attractant (Bono & Villu Maricq, 2005). One can predict the stereotyped behaviour with certainty (Bono & Villu Maricq, 2005), one might know explicitly that this is a simple organism with only 302 neurons (White, Southgate, Thomson, & Brenner, 1986), and know the whole genome (Yook et al., 2012). Even though in practice it is difficult to predict the behaviour of model organisms (Maye, Hsieh, Sugihara, & Brembs, 2007), in theory, every step of the behaviour taken by itself could potentially be explained mechanistically. When watching such single cellular organism move towards a pipette, one can speculate that just as with Heider and Simmel stimuli (Scholl & Tremoulet, 2000), most people, even scientists, would probably spontaneously attribute mental states, agency and intentionality to such organisms. Cues that trigger the detection of animacy include non-Newtonian motion and sudden changes in motion direction and speed, and the perception of animacy from such cues is so automatic and irresistible that we even attribute animacy to two-dimensional shapes such as triangles moving around on a computer screen. What is the reason for this? One potential explanation, which I proposed in Chapter 5, is that this is a heuristic to predict biological agents behaviour, which is difficult (too complex) to model. Animate motion is less predictable than inanimate motion because it is more complex and nonlinear, and because self-propulsion relies on hidden internal (intentional) causes. Due to the inherent unpredictability, viewing of animate motion compared with inanimate movement should result in larger and more informative sensory prediction error signal (i.e. a mismatch between the model's prediction and the sensory input). In chapter 5, we investigated effective connectivity between brain regions involved in the perception of animacy.

Animate motion (for example, the motion of animals and other biological agents) is often more complex than mechanical motion (for example, the motion of billiard balls or rockets) and thus it should be more difficult for the brain to predict animate motion. The resulting errors of prediction should result in a larger error signal from sensory brain regions, resulting in activity in regions that predict the motion trajectories in animate motion. In chapter 5, we tested this hypothesis by investigating effective connectivity in a large-scale fMRI dataset from the Human Connectome Project. Participants viewed animations of triangles that were either animate, moving intentionally, or inanimate (moving in a mechanical way). We provided evidence for this, by using a novel technique allowing us to compare many mathematical models' fit.

Predictions about animate motion – relative to mechanical motion –increased signal passing from lower level sensory area MT+/V5, which is responsive to all motion, to higher-order posterior superior temporal sulcus, which is selectively activated by animate motion. Specifically, we found that forward connectivity from V5 to pSTS increased, and inhibitory self-connection in the pSTS decreased, when viewing intentional motion versus inanimate motion. This may suggest that animate motion prediction error may underlie enhanced attention and the phenomenological states of agency perception. Data and analysis programs have been made open access online to enable other researchers to replicate and extend these findings.

In line with this, it has recently been shown – using the violation-of-expectation-method (for a critical review, see (Munakata, 2000)) – that infants as young as 8 months old are surprised, if they find out that biological agents that move in a self-propelled way are hollow inside (Setoh, Wu, Baillargeon, & Gelman, 2013). This suggests that infants expect there to be hidden (intentional) causes of the self-propelled biological movement. Some have argued that mentalizing and the predictions of intentions use similar predictive coding principles as are involved in vision (Frith & Frith, 2012; Kilner, Friston, & Frith, 2007). First, based on an estimated intention, an agent's behaviour is predicted, which results in a prediction error on the basis of which the estimated intention (model of the agent) is updated (Frith & Frith, 2012). Potentially, as eye tracking studies suggest more frequent fixations on animate motion vs. inanimate motion, these predictions might have to be updated more rapidly (Klein, Zwickel, Prinz, & Frith, 2009). In other words, the predictive power of a model about an animate object like a human breaks down very quickly (due to nonlinearities): when someone reaches for a bottle one can infer that person is thirsty and will drink soon. However, one does not know where someone will be 30 minutes later based on an initial look, and one's model of another person needs to be constantly updated in order to predict their behaviour. In contrast, one's model of nonanimate objects do not have break down as fast as they are more linear. For instance, my 30-minute-old model of the moon and my prediction of its position will be much more accurate than my 30-minute old model of an animate agent that I have not updated.

6.5.1. Hierachical predictive coding of motion in the visual system

I would like to outline future directions for potential experiments that follow from a theory that is based on the results in chapter 5. In chapter 5, I argued that animate motion trajectories are more difficult to model and predict based on an internal generative model in the brain. When watching animate vs. inanimate motion higher level brain pSTS was activated mostly by sensory prediction errors coming from V5. In other words, V5 could not predict all the sensory input and sends the prediction errors forward to higher level areas and receives backward predictions from higher level areas like pSTS. I propose that one could break down the lower level properties that these different motion trajectories possess and model them mathematically. I hypothesize that the visual system could be hierarchically activated by stimuli that can be modelled by position and its derivatives. In this model, V1, lowest in the visual hierarchy is activated by position (visual stimulation that is not moving), V5 is activated by the derivative of position, which is velocity or motion, and higher-level areas like pSTS are activated by higher order derivatives of position like acceleration and jerk.

Imagine a dot that if displayed will activate V1. To predict where a moving dot will be next, one needs to compute the derivative of position, which is velocity or the first derivative (or rate of change) of position. In hierarchical predictive coding frameworks, higher level areas predict the responses in lower level areas. Thus, in order to predict the responses in V1, a higher level area such as V5, has to compute and predict its activity. Modelling the position derivative helps to predict the future position (if one wants to know where something will be in space at a later time than current time one needs to know its speed and direction). Crucially, the higher order the derivative, the slower the time scale is fluctuating (one needs to observe at least two points of data to know about whether something is moving and it takes longer to gather two data points than one). If the dot is not moving, the derivative of position (velocity) does not need to be computed and then V5 will not activate (and there will not be any prediction error from V1). In the data presented in chapter 5, all motion activated V5, but V5 was not significantly more activated in the animate condition, while pSTS was. This suggests that the more complex motion trajectories might selectively activate higher-level areas.

The derivate of velocity (and the second order derivative of position) is acceleration. This is the rate of change of velocity and can predict the velocity in an even slower time scale (at least three data points are needed to calculate acceleration). Next, jerk is the rate of change of acceleration with respect to time (the second order derivative of velocity and the third order derivative of position). In chapter 5, our animate motion notably differs from the inanimate motion with respect to their acceleration and jerk profiles. Animate motion has a different jerk profile compared with inanimate motion (animate motion has what has been called minimum jerk velocity (Cook, Saygin, Swain, & Blakemore, 2009)). In my study, these differences in acceleration and/or jerk might be components underlying the animate motion that activated pSTS. Future research could strategically manipulate the position, velocity, acceleration, jerk and jounce (sometimes called ‘snap’ – the rate of change of jerk with respect to time) of dots moving and see these position derivatives have neural correlates in areas that are ordered hierarchically in the sense that they activate increasingly higher-level areas along the rostral caudal axis (V1, V5, pSTS). This would constitute a direct test of whether the visual cortex is organized (roughly) along the rostro-caudal gradient, going from primary sensory to association to higher level areas (Kiebel, Daunizeau, & Friston, 2008). This hierarchy of position derivatives would also mirror the temporal structure of an agent's environment: stimuli that fluctuate faster, like position and velocity, activate the more sensory areas (V1 and V5), whereas increasingly slower fluctuation (like acceleration, jerk and jounce) activate ever higher level areas like pSTS (Kiebel et al., 2008). An fMRI study by Cook et al. (Cook, Press, Saygin, Kilner, & Blakemore, in prep) varied the velocity profiles parametrically going from 'smoothly moving' minimum jerk (looking more biological) to uniform acceleration maximum jerk. Subjective judgment of whether the source of motion was biological activated pSTS. It might be that pSTS serves as a biological motion detection area by way of being activated by the sensory prediction errors due to the complexity that is typical of biological motion. This might then cause other brain areas to respond correspondingly: for instance, some studies have shown increased saccades in response to animate (complex) motion (Klein et al., 2009); this can be interpreted as collecting more sensory data to predict the motion and to decrease the sensory prediction error (Friston, Adams, Perrinet, & Breakspear, 2012).

6.6. General limitations

6.6.1. Limitations of having participants of only one gender (Chapter 2 - 4)

In chapters 2 to 4, we only tested female participants and this raises the concern of whether the findings presented here are generalizable to other genders.

In chapter 2, the reason for this were pragmatic: in order to control for gender effects and so that everyone would 'play' with players of her own gender only female participants were recruited. It would be interesting to see whether our results replicate in a sample of other genders and how interactions between different genders would evolve. One recent study finds some limited influence for gender information on trust behaviour (Bonein & Serra, 2009), but a recent review concludes that the findings on gender effects in trust games are mixed (Ergun, García-Muñoz, & Rivas, 2012).

In chapter 3 and 4, we also used only female participants. This was to increase the signal to noise ratio, as there are sex differences in brain structure and function (Cahill, 2006) and it could potentially be important to decrease the variability and thus the noise in brain structure in order to pick up small, but robust effects. Though counterintuitive, this line of argument does not necessarily imply that the findings are not generalizable. One might imagine decreasing the noise of a sample by only collecting data from people from a certain height range and finding a small and robust effect. This effect might then be generalizable to a population of people with different heights. To be sure, the findings in this thesis should be replicated with participants from other genders.

6.6.2. Interhemispheric integration (Chapter 4 & 5)

The dynamic causal models examining effective connectivity in the extant research were all in one hemisphere. Even though, using the high quality data from the HCP, we were able to replicate our models in both hemispheres and the parameter estimates of the models in the two hemispheres were highly correlated, it would be interesting to use interhemispheric models in future research. We did not venture in this direction as the mechanisms underlying interhemispheric integration remain poorly understood (Stephan, Marshall, Penny, Friston, & Fink, 2007) and there are few DCMs that explore interhemispheric integration (for an example, see Stephan, Marshall, Penny, Friston, & Fink, 2007).

6.6.3. Dynamic Causal modelling (Chapter 4 & 5)

Finally, the relative strengths and weaknesses of dynamic causal modelling in comparison to other effective connectivity approaches such as Granger causality (GC) are manifold and have been discussed elsewhere in detail (Friston, Moran, & Seth, 2013). Most importantly for our data here is that DCM has been argued to be more suited than for fMRI data, because DCM models haemodynamic variations explicitly and uses a generative model to explain the hidden neuronal states in the data (Friston et al., 2013).

6.6.4. Differences in methodology and statistics (Chapter 3-5)

Unfortunately, pragmatic concerns (i.e. not having very high statistical power due to not having a very large number of participants) forced us to use slightly different methodological and statistical approaches in the different chapters of this thesis. For instance, both the search radii for VOI extraction and the radii of the actual VOI spheres that we extracted for DCM analysis were different in the chapters 4 and 5. The setting of those radii is a trade-off between capturing smaller, less robust effects (more diffuse activations) and discarding fewer participants that do not show the activation vs. having higher anatomical accuracy (less intersubject variability). In chapter 4, we had lower statistical power than in chapter 5, where we had more participants, and so in chapter 4, we had to lower the statistical threshold and increase the VOI extraction radii. In chapter 5, we could afford higher anatomical accuracy due to high statistical power. Additionally, in chapter 5, the regions analysed were closer together, and so we chose a smaller radius in order to not have any overlap between the VOIs. Finally, the process of extracting larger VOIs with many voxels (but not the process of inverting the resulting DCMs) is quite computationally expensive, because as the VOI spheres increases in diameter, the number of voxels extracted increases exponentially.

We also used different approaches to statistical thresholding in the different GLM analyses of regional activations in the different chapters. Specifically, in chapter 3, we used cluster-level FWE thresholding; whereas in chapter 5, we used voxel-level FWE corrected thresholding. This was due to lower statistical power in the study that chapter 3 is based on. Cluster level thresholding has higher sensitivity than voxel based corrections (fewer type II errors) and is recommended for studies with a moderate number of participants and moderate effect sizes (Woo, Krishnan, & Wager, 2014). If the initial p-value is set to be sufficiently conservative and the cluster extent threshold is set correctly, according to the level of smoothness in the images, then cluster-level as opposed to voxel-level thresholding has no inherently greater risk for type I errors (Woo et al., 2014). Some people advocate liberal initial p-values and an extent threshold of 10 or 20 voxels (Lieberman & Cunningham, 2009), however, this has recently been criticised too liberal (p-value) and arbitrary (cluster-extent) (Woo et al., 2014). In chapter 3, we used an initial p value of $p< 0.001$, as recommended, and a cluster extent threshold of 82 voxels as calculated on our own images (Woo et al., 2014), in order to follow best practise. Finally, even correctly controlled, large cluster-corrected activations are frequently misinterpreted: it is sometimes implied that there is signal everywhere in the cluster, whereas in actuality, one cannot make inference about all the regions in a large cluster, but only that within the large cluster there is significant activation (signal) somewhere and the rest is likely mostly noise (Woo et al., 2014). Thus, in our study reported in chapter 3, we cannot be sure about whether e.g. the STG/STS subpeak with a small z-score of 3.9, which is within a larger STG/STS cluster, is actually active.

Ideally, one would do a priori power calculations, then test as a many participants as necessary to achieve high standardized power, in order to allow for conservative standardized thresholds and also make unthresholded activation maps available for meta-analyses (as opposed to lowering the threshold). However, power analysis is uncommon and difficult for fMRI studies (see Mumford, 2012 for a discussion), because effect sizes are difficult to estimate for new paradigms and achieving high power is mostly possible by increasing the number of participants, which is costly. Because some of the results presented here are somewhat weak, they must be treated with caution and replicated in future studies, as they could be due to type I errors (i.e. false positives).

6.7. Conclusion and implications

I will now draw a few more general conclusions that are relevant to the field of social neuroscience, but these conclusions do not specifically apply to the work presented in this thesis and are of speculative nature. Social cognitive neuroscience is an important field of study for many reasons. Apart from the relevance to inform more basic research, the most applied reason is perhaps that it is very clinically relevant, since in almost every condition that affects mental health, social functioning is impaired (Adolphs, 2010). Social functioning can be seen as one of the most high-level skills humans perform and for this reason it might be prone to break down disproportionately often. Potentially, responses to social neuroscience tasks might be used as biomarkers of mental health problems or even to predict mental health problems in advance. Some recent research has focused on task-free measures, such as rsfMRI, and machine learning approaches to predict clinical brain states (for a review, see Lee, Smyser, & Shimony, 2013). More recently, some research has focused on task based fMRI to predict clinical brain states, with the aim of elucidating the underlying neuronal mechanisms (see for instance, Brodersen et al., 2014). Moreover, while some promising neuroimaging research has attempted to predict disease status, neuroimaging has generally had limited clinical reliability (for a discussion see, Borgwardt, Radua, Mechelli, & Fusar-Poli, 2012). Another reason for this might be that some studies use machine learning approaches, which have high data dimensionality and low sample sizes, and algorithms often pick up on uninformative features of the data, resulting in low generalizability (see Brodersen et al., 2014 for a discussion).

Furthermore, many mental health problems, such as schizophrenia have been suggested to be problems of aberrant functional connectivity between distinct brain regions (Pettersson-Yeo, Allen, Benetti, McGuire, & Mechelli, 2011). For this reason, the connectivity might be the most informative features of the data and we can use it to predict clinical brain states. For instance, there is some initial research showing that models of connectivity (DCM) can outperform traditional discrimative approaches such as support vector machines in predicting clinical brain states (Brodersen et al., 2011). Another advantage of using connectivity measures to predict disease is that then the connectivity parameters can be interpreted mechanistically, in order to not only advance our theories about the ætiology of disease, but also predict disease progression, dissect spectrum disorders, and eventually adjust medications (Brodersen et al., 2011).

Another applied reason to study social neuroscience is that it helps us to understand collective behaviour and collaboration, which might improve public policy making, design of social computer interfaces and robotics (Adolphs, 2010).

Finally, social neuroscience studies what are arguably the most ‘sacred’ computations the human mind is able to make, such as attraction, stereotypes, and moral judgments. Elucidating the mechanistic (neural) basis of such high-level social cognitions, might undermine the public’s dualistic beliefs, resulting in positive downstream consequences for society (Greene & Cohen, 2004). For instance, it might be that due to the notion of the ‘bad soul’, legal systems around the world are influenced by retributivism, i.e. sometimes punish for punishment’s sake. It has been argued that, if we were to view the high level computations that the mind makes more mechanistically, the public might intuit that there are causal reasons for bad behaviour, and the legal system might switch to a more consequentialist system, that aims to benefit the greater good (Greene & Cohen, 2004).

7. References

7.1. Chapter 1

Apperly, I. A. (2012). What is “theory of mind”? Concepts, cognitive processes and individual differences. *Quarterly Journal of Experimental Psychology (2006)*, *65*(5), 825–839. doi:10.1080/17470218.2012.676055

Apperly, I. A., Carroll, D. J., Samson, D., Humphreys, G. W., Qureshi, A., & Moffitt, G. (2010a). Why are there limits on theory of mind use? Evidence from adults’ ability to follow instructions from an ignorant speaker. *Quarterly Journal of Experimental Psychology (2006)*, *63*(6), 1201–1217. doi:10.1080/17470210903281582

Apperly, I. A., Carroll, D. J., Samson, D., Humphreys, G. W., Qureshi, A., & Moffitt, G. (2010b). Why are there limits on theory of mind use? Evidence from adults’ ability to follow instructions from an ignorant speaker. *Quarterly Journal of Experimental Psychology*, *63*(6), 1201–1217. doi:10.1080/17470210903281582

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*(4489), 1390–1396. doi:10.1126/science.7466396

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind” ? *Cognition*, *21*(1), 37–46. doi:10.1016/0010-0277(85)90022-8

Blakemore, S.-J. (2008). The social brain in adolescence. *Nature Reviews. Neuroscience*, *9*(4), 267–277. doi:10.1038/nrn2353

Blakemore, S.-J., & Robbins, T. W. (2012). Decision-making in the adolescent brain. *Nature Neuroscience*, *15*(9), 1184–1191. doi:10.1038/nn.3177

Brown, B. B. (2004). Adolescents’ relationships with peers. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (pp. 363–394). Hoboken, NJ: Wiley.

Brown, E. C., & Brüne, M. (2012). The role of prediction in social neuroscience. *Frontiers in Human Neuroscience*, *6*, 147. doi:10.3389/fnhum.2012.00147

Brown, H., & Friston, K. (2012). Free-Energy and Illusions: The Cornsweet Effect. *Frontiers in Psychology*, *3*. doi:10.3389/fpsyg.2012.00043

Burnett, S., Bird, G., Moll, J., Frith, C., & Blakemore, S.-J. (2009). Development during adolescence of the neural processing of social emotion. *Journal of Cognitive Neuroscience*, *21*(9), 1736–1750. doi:10.1162/jocn.2009.21121

Burnett, S., & Blakemore, S.-J. (2009). Functional connectivity during a social emotion task in adolescents and in adults. *The European Journal of Neuroscience*, *29*(6), 1294–1301.

Burnett, S., Sebastian, C., Cohen Kadosh, K., & Blakemore, S.-J. (2011). The social brain in adolescence: evidence from functional magnetic resonance imaging and behavioural studies. *Neuroscience and Biobehavioral Reviews*, *35*(8), 1654–1664. doi:10.1016/j.neubiorev.2010.10.011

Cacioppo, J. T. (2002). Social neuroscience: Understanding the pieces fosters understanding the whole and vice versa. *American Psychologist*, *57*(11), 819–831. doi:10.1037/0003-066X.57.11.819

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, *12*(5), 187–192. doi:10.1016/j.tics.2008.02.010

Camerer, C. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, *63*(4), i–vi, 1–143.

Clark. (2012). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences.*, *Sec. 3*, 1–86.

Crone, E. A. (2009). Executive functions in adolescence: inferences from brain and behavior. *Developmental Science*, *12*(6), 825–830.

Crone, E. A., & Ridderinkhof, K. R. (2010). The developing brain: From theory to neuroimaging and back. *Developmental Cognitive Neuroscience*, *1*(2), 101–109. doi:10.1016/j.dcn.2010.12.001

Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience*, *16*(6), 763–770. doi:10.1038/nn.3381

Dennett, D. C. (1981). *Brainstorms: Philosophical Essays on Mind and Psychology*. MIT Press.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale*, *91*(1), 176–180.

Dumontheil, I., Apperly, I. A., & Blakemore, S. J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, *13*, 331–338.

Dumontheil, I., & Blakemore, S.-J. (n.d.). Social cognition and abstract thought in adolescence: The role of structural and functional development in rostral prefrontal cortex. *British Journal of Educational Psychology*.

Dunne, S., & O’Doherty, J. P. (2013). Insights from the application of computational neuroimaging to social neuroscience. *Current Opinion in Neurobiology*, *23*(3), 387–392. doi:10.1016/j.conb.2013.02.007

Egner, T. (2011). Surprise! A unifying model of dorsal anterior cingulate function? *Nature Neuroscience*, *14*(10), 1219–1220. doi:10.1038/nn.2932

Egner, T., Monti, J. M., & Summerfield, C. (2010). Expectation and Surprise Determine Neural Population Responses in the Ventral Visual Stream. *The Journal of Neuroscience*, *30*(49), 16601–16608. doi:10.1523/JNEUROSCI.2770-10.2010

Eisenberg, N., & Morris, A. S. (2004). Moral cognitions and prosocial responding in adolescence. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (pp. 155–188). Hoboken, NJ: Wiley.

Fehr, E., & Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends in Cognitive Sciences*, *11*(10), 419–427. doi:10.1016/j.tics.2007.09.002

Fletcher, P. C., Happé, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., & 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.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Essen, D. C. V., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(27), 9673–9678. doi:10.1073/pnas.0504136102

Friston, K. (2002a). Beyond phrenology: what can neuroimaging tell us about distributed circuitry? *Annual Review of Neuroscience*, *25*, 221–250. doi:10.1146/annurev.neuro.25.112701.142846

Friston, K. (2002b). Functional integration and inference in the brain. *Progress in Neurobiology*, *68*(2), 113–143. doi:10.1016/S0301-0082(02)00076-X

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews. Neuroscience*, *11*(2), 127–138. doi:10.1038/nrn2787

Frith, U., & Frith, C. (2001). The Biological Basis of Social Interaction. *Current Directions in Psychological Science*, *10*(5), 151–155. doi:10.1111/1467-8721.00137

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

Fry, D. P., & Söderberg, P. (2013). Lethal Aggression in Mobile Forager Bands and Implications for the Origins of War. *Science*, *341*(6143), 270–273. doi:10.1126/science.1235675

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of “theory of mind.” *Trends in Cognitive Sciences*, *7*(2), 77–83.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., … Rapoport, J. L. (1999). Brain development during childhood and adolescence: a longitudinal MRI study. *Nature Neuroscience*, *2*(10), 861–863. doi:10.1038/13158

Ginesu, G., Pintus, M., & Giusto, D. D. (2012). Objective assessment of the WebP image coding algorithm. *Signal Processing: Image Communication*, *27*(8), 867–874. doi:10.1016/j.image.2012.01.011

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., … Thompson, P. M. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(21), 8174–8179. doi:10.1073/pnas.0402680101

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*(1), 14–23. doi:10.1016/j.tics.2005.11.006

Happé, F., Ehlers, S., Fletcher, P., Frith, U., Johansson, M., Gillberg, C., … Frith, C. (1996). “Theory of mind” in the brain. Evidence from a PET scan study of Asperger syndrome. *Neuroreport*, *8*(1), 197–201.

Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour*, *61*(1), 139–151. doi:10.1006/anbe.2000.1518

Harris, J. J., Reynell, C., & Attwell, D. (2011). The physiology of developmental changes in BOLD functional imaging signals. *Developmental Cognitive Neuroscience*, *1*(3), 199–216. doi:10.1016/j.dcn.2011.04.001

Iacoboni, M., & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience*, *7*(12), 942–951. doi:10.1038/nrn2024

Iaria, G., Fox, C. J., Waite, C. T., Aharon, I., & Barton, J. J. S. (2008). The contribution of the fusiform gyrus and superior temporal sulcus in processing facial attractiveness: neuropsychological and neuroimaging evidence. *Neuroscience*, *155*(2), 409–422. doi:10.1016/j.neuroscience.2008.05.046

Jenkins, A. C., Macrae, C. N., & Mitchell, J. P. (2008). Repetition suppression of ventromedial prefrontal activity during judgments of self and others. *Proceedings of the National Academy of Sciences*, *105*(11), 4507–4512. doi:10.1073/pnas.0708785105

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: the role of mutual knowledge in comprehension. *Psychological Science : A Journal of the American Psychological Society / APS*, *11*(1), 32–38.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*(1), 25–41.

Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A Hierarchy of Time-Scales and the Brain. *PLoS Comput Biol*, *4*(11), e1000209. doi:10.1371/journal.pcbi.1000209

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. doi:10.1126/science.1108062

Kovács, G., Kaiser, D., Kaliukhovich, D. A., Vidnyánszky, Z., & Vogels, R. (2013). Repetition probability does not affect fMRI repetition suppression for objects. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(23), 9805–9812. doi:10.1523/JNEUROSCI.3423-12.2013

Liebenberg, L. (2008). The relevance of persistence hunting to human evolution. *Journal of Human Evolution*, *55*(6), 1156–1159. doi:10.1016/j.jhevol.2008.07.004

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, *453*(7197), 869–878. doi:10.1038/nature06976

Luna, B., Padmanabhan, A., & O'Hearn, K. (2010). What has fMRI told us about the development of cognitive control through adolescence? *Brain and Cognition*, *72*(1), 101–113. doi:10.1016/j.bandc.2009.08.005

Mar, R. A. (2011). The neural bases of social cognition and story comprehension. *Annual Review of Psychology*, *62*, 103–134. doi:10.1146/annurev-psych-120709-145406

Mills, K. L., Lalonde, F., Clasen, L. S., Giedd, J. N., & Blakemore, S.-J. (2012). Developmental changes in the structure of the social brain in late childhood and adolescence. *Social Cognitive and Affective Neuroscience*. doi:10.1093/scan/nss113

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-Neuron Responses in Humans during Execution and Observation of Actions. *Current Biology*, *20*(8), 750–756. doi:10.1016/j.cub.2010.02.045

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, *87*(24), 9868–9872.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-Month-Old Infants Understand False Beliefs? *Science*, *308*(5719), 255 –258. doi:10.1126/science.1107621

Optical illusion. (2013, September 4). In *Wikipedia, the free encyclopedia*. Retrieved from http://en.wikipedia.org/w/index.php?title=Optical_illusion&oldid=571489627

Passingham, R. E., Rowe, J. B., & Sakai, K. (2013). Has brain imaging discovered anything new about how the brain works? *NeuroImage*, *66*, 142–150. doi:10.1016/j.neuroimage.2012.10.079

Peterson, C. C., Slaughter, V., Peterson, J., & Premack, D. (2013). Children with autism can track others' beliefs in a competitive game. *Developmental Science*, *16*(3), 443–450. doi:10.1111/desc.12040

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(04), 515–526. doi:10.1017/S0140525X00076512

Przyrembel, M., Smallwood, J., & Singer, T. (2012). Illuminating the dark matter of social neuroscience: Considering the problem of social interaction from philosophical, psychological, and neuroscientific perspectives. *Frontiers in Human Neuroscience*, *6*, 190. doi:10.3389/fnhum.2012.00190

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, *9*(7), 545–556. doi:10.1038/nrn2357

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology*, *57*, 199–226. doi:10.1146/annurev.psych.57.102904.190208

Risko, E. F., Laidlaw, K. E. W., Freeth, M., Foulsham, T., & Kingstone, A. (2012). Social attention with real versus reel stimuli: toward an empirical approach to concerns about ecological validity. *Frontiers in Human Neuroscience*, *6*, 143. doi:10.3389/fnhum.2012.00143

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, *2*(9), 661–670. doi:10.1038/35090060

Rushworth, M. F., Mars, R. B., & Sallet, J. (n.d.). Are there specialized circuits for social cognition and are they unique to humans? *Current Opinion in Neurobiology*, (0). doi:10.1016/j.conb.2012.11.013

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology. Human Perception and Performance*, *36*(5), 1255–1266. doi:10.1037/a0018729

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, *16*(2), 235–239.

Schilbach, L. (2010). A second-person approach to other minds. *Nat Rev Neurosci*, *11*(6), 449. doi:10.1038/nrn2805-c1

Sengupta, B., Stemmler, M. B., & Friston, K. J. (2013). Information and Efficiency in the Nervous System—A Synthesis. *PLoS Comput Biol*, *9*(7), e1003157. doi:10.1371/journal.pcbi.1003157

Shaw, P., Kabani, N. J., Lerch, J. P., Eckstrand, K., Lenroot, R., Gogtay, N., … Wise, S. P. (2008). Neurodevelopmental trajectories of the human cerebral cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *28*(14), 3586–3594. doi:10.1523/JNEUROSCI.5309-07.2008

Sowell, E. R., Thompson, P. M., Holmes, C. J., Batth, R., Jernigan, T. L., & Toga, A. W. (1999). Localizing age-related changes in brain structure between childhood and adolescence using statistical parametric mapping. *NeuroImage*, *9*(6 Pt 1), 587–597. doi:10.1006/nimg.1999.0436

Stephan, K. E., & Friston, K. J. (2010). Analyzing effective connectivity with functional magnetic resonance imaging. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(3), 446–459. doi:10.1002/wcs.58

Stephan, K. E., Harrison, L. M., Kiebel, S. J., David, O., Penny, W. D., & Friston, K. J. (2007). Dynamic causal models of neural system dynamics:current state and future extensions. *Journal of Biosciences*, *32*(1), 129–144.

Stephan, K. E., Penny, W., Moran, R. J., den Ouden, H. E. M., Daunizeau, J., & Friston, K. (2010). Ten simple rules for dynamic causal modeling. *NeuroImage*, *49*(4), 3099–3109. doi:10.1016/j.neuroimage.2009.11.015

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M.-M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11*(9), 1004–1006. doi:10.1038/nn.2163

Van Overwalle, F., Baetens, K., Mariën, P., & Vandekerckhove, M. (2013). Social Cognition and the Cerebellum: A Meta-analysis of over 350 fMRI studies. *NeuroImage*. doi:10.1016/j.neuroimage.2013.09.033

Ward, J. (2012). *The student's guide to social neuroscience*. New York: Psychology Press.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-Analysis of Theory-of-Mind Development: The Truth about False Belief. *Child Development*, *72*(3), 655–684. doi:10.1111/1467-8624.00304

Williams, K. D. (2007). Ostracism. *Annual Review of Psychology*, 58, 425–452. doi:10.1146/annurev.psych.58.110405.085641

Winking, J., & Mizer, N. (2013). Natural-field dictator game shows no altruistic giving. *Evolution and Human Behavior*, *34*(4), 288–293. doi:10.1016/j.evolhumbehav.2013.04.002

Yirmiya, N., Erel, O., Shaked, M., & Solomonica-Levi, D. (1998). Meta-analyses comparing theory of mind abilities of individuals with autism, individuals with mental retardation, and normally developing individuals. *Psychological Bulletin*, *124*(3), 283–307.

7.2. Chapter 2

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the Structure of Interpersonal Closeness. *Journal of Personality and Social Psychology*, 63(4), 596–612.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavio*r, 10(1), 122–142. doi:06/game.1995.1027

Bracht, J., & Feltovich, N. (2009). Whatever you say, your reputation precedes you: Observation and cheap talk in the trust game. *Journal of Public Economics*, 93(9-10), 1036–1044. doi:16/j.jpubeco.2009.06.004

Carter-Sowell, A., Chen, Z., & Williams, K. (2008). Ostracism increases social susceptibility. *Social Influence*, 3(3), 143–153. doi:10.1080/15534510802204868

Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nat Neurosci*, 8(11), 1611–1618. doi:10.1038/nn1575

Eisenberger, N. I., Gable, S. L., & Lieberman, M. D. (2007). Functional magnetic resonance imaging responses relate to differences in real-world social experience. *Emotion* (Washington, D.C.), 7(4), 745–754. doi:10.1037/1528-3542.7.4.745

Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does Rejection Hurt? An fMRI Study of Social Exclusion. *Science*, 302(5643), 290 –292. doi:10.1126/science.1089134

Goto, S. G. (1996). To trust or not to trust: Situational and dispositional determinants. *Social Behavior and Personality: an international journal*, 24, 119–131. doi:10.2224/sbp.1996.24.2.119

Heinrichs, M., von Dawans, B., & Domes, G. (2009). Oxytocin, vasopressin, and human social behavior. *Frontiers in Neuroendocrinology*, 30(4), 548–557. doi:10.1016/j.yfrne.2009.05.005

Hillebrandt, H., Sebastian, C., & Blakemore, S.-J. (2011). Experimentally induced social inclusion influences behavior on trust games. *Cognitive Neuroscience,* 2(1), 27. doi:10.1080/17588928.2010.515020

Johnson, N. D., & Mislin, A. (n.d.). Cultures of Kindness: A Meta-Analysis of Trust Game Experiments. Retrieved from *Social Science Research Network (SSRN):* http://ssrn.com/abstract=1315325

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. doi:10.1126/science.1108062

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature*, 435(7042), 673–676. doi:10.1038/nature03701

Kross, E., Egner, T., Ochsner, K., Hirsch, J., & Downey, G. (2007). Neural dynamics of rejection sensitivity. *Journal of Cognitive Neuroscience*, 19(6), 945–956. doi:10.1162/jocn.2007.19.6.945

Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., … Grafman, J. (2007). Neural correlates of trust. P*roceedings of the National Academy of Sciences of the United States of America*, 104(50), 20084–20089. doi:10.1073/pnas.0710103104

Lakin, J. L., Chartrand, T. L., & Arkin, R. M. (2008). I am too just like you: nonconscious mimicry as an automatic behavioral response to social exclusion. *Psychological Science,* 19(8), 816–822. doi:10.1111/j.1467-9280.2008.02162.x

Lieberman, M. D., & Eisenberger, N. I. (2009). Neuroscience. Pains and pleasures of social life. *Science* (New York, N.Y.), 323(5916), 890–891. doi:10.1126/science.1170008

Majolo, B., Ames, K., Brumpton, R., Garratt, R., Hall, K., & Wilson, N. (2006). Human friendship favours cooperation in the Iterated Prisoner's Dilemma. *Behaviour*, 143(11), 1383–1395. doi:10.1163/156853906778987506

Maner, J. K., DeWall, C. N., Baumeister, R. F., & Schaller, M. (2007). Does social exclusion motivate interpersonal reconnection? Resolving the "porcupine problem." J*ournal of Personality and Social Psychology*, 92(1), 42–55. doi:10.1037/0022-3514.92.1.42

Masten, C. L., Eisenberger, N. I., Borofsky, L. A., Pfeifer, J. H., McNealy, K., Mazziotta, J. C., & Dapretto, M. (2009). Neural correlates of social exclusion during adolescence: understanding the distress of peer rejection. *Social Cognitive and Affective Neuroscience*, 4(2), 143–157. doi:10.1093/scan/nsp007

Pennisi, E. (2005). How Did Cooperative Behavior Evolve? *Science*, 309(5731), 93. doi:10.1126/science.309.5731.93

Somerville, L. H., Heatherton, T. F., & Kelley, W. M. (2006). Anterior cingulate cortex responds differentially to expectancy violation and social rejection. *Nature Neuroscience*, 9(8), 1007–1008. doi:10.1038/nn1728

Todorov, A., Harris, L. T., & Fiske, S. T. (2006). Toward socially inspired social neuroscience. *Brain Research*, 1079(1), 76–85. doi:10.1016/j.brainres.2005.12.114

West, S. A., El Mouden, C., & Gardner, A. (2011). Sixteen common misconceptions about the evolution of cooperation in humans. *Evolution and Human Behavior*, 32(4), 231–262. doi:16/j.evolhumbehav.2010.08.001

Williams, K D, Cheung, C. K., & Choi, W. (2000). Cyberostracism: effects of being ignored over the Internet. *Journal of Personality and Social Psychology*, 79(5), 748–762.

Williams, Kipling D. (2007). Ostracism. *Annual Review of Psychology*, 58, 425–452. doi:10.1146/annurev.psych.58.110405.085641

Williams, Kipling D., & Nida, S. A. (2011). Ostracism Consequences and Coping. *Current Directions in Psychological Science*, 20(2), 71–75. doi:10.1177/0963721411402480

Williams, Kipling D., & Sommer, K. L. (1997). Social Ostracism by Coworkers: Does Rejection Lead to Loafing or Compensation? *Personality and Social Psychology Bulletin*, 23(7), 693 –706. doi:10.1177/0146167297237003

7.3. Chapter 3

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

Apperly, I. A. (2011). *Mindreaders: The cognitive basis of "Theory of mind."* Psychology Press, Hove, UK.

Apperly, I. A., Carroll, D. J., Samson, D., Humphreys, G. W., Qureshi, A., & Moffitt, G. (2010). Why are there limits on theory of mind use? Evidence from adults' ability to follow instructions from an ignorant speaker. *Quarterly Journal of Experimental Psychology*, *63*, 1201-1217.

Apperly, I. A., Samson, D., Chiavarino, C., & Humphreys, G. W. (2004). Frontal and temporo-parietal lobe contributions to theory of mind: Neuropsychological evidence from a false-belief task with reduced language and executive demands. *Journal of Cognitive Neuroscience*, *16*, 1773-1784.

Astle, D. E., & Scerif, G. (2009). Using developmental cognitive neuroscience to study behavioral and attentional control. *Developmental psychobiology*, *51*, 107-118.

Back, E., & Apperly, I. A. (2010). Two sources of evidence on the non-automaticity of true and false belief ascription. *Cognition*, *115*, 54-70.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1986). Mechanical, behavioural and intentional understanding of picture stories in autistic children. *British Journal of Developmental Psychology*, *4*, 113-125.

Beck, D. M., & Kastner, S. (2009). Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vision Research*, *49*, 1154-1165.

Blakemore, S. J. (2008). The social brain in adolescence. *Nature Reviews Neuroscience*, *9*, 267-277.

Blakemore, S.-J., Burnett, S., & Dahl, R. E. (2010). The role of puberty in the developing adolescent brain. *Human Brain Mapping*, *31*, 926-33.

Blakemore, S. J., den Ouden, H., Choudhury, S., & Frith, C. (2007). Adolescent development of the neural circuitry for thinking about intentions. *Social Cognitive and Affective Neuroscience*, *2*, 130-139.

Booth, J. R., Burman, D. D., Meyer, J. R., Lei, Z., Trommer, B. L., Davenport, N. D., Li, W., et al. (2003). Neural development of selective attention and response inhibition. *NeuroImage*, *20*, 737-751.

Brothers, L. (1990). The social brain: A project for integrating primate behavior and neurophysiology in a new domain. *Concepts in Neuroscience*, *1*, 27-51.

Brown, B. B. (2004). Adolescents' relationships with peers. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of Adolescent Psychology* (pp. 363-394). Hoboken, NJ: Wiley.

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

Burgess, P. W., Gilbert, S J, & Dumontheil, I. (2007). Function and localization within rostral prefrontal cortex (area 10). *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, *362*, 887-899.

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

Burnett, S., Bird, G., Moll, J., Frith, C., & Blakemore, S.-J. (2008). Development during adolescence of the neural processing of social emotion. *Journal of Cognitive Neuroscience, 21,* 1736-1750.

Burnett, S., Sebastian, C., Cohen-Kadosh, K., & Blakemore, S.-J. (2011). The social brain in adolescence: Evidence from functional magnetic resonance imaging and behavioural studies. *Neuroscience and Biobehavioral Reviews, 35,* 1654-1664.

Buschman, T. J., & Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science*, *315*, 1860-1862.

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

Chandrasekaran, C., & Ghazanfar, A. A. (2009). Different neural frequency bands integrate faces and voices differently in the superior temporal sulcus. *Journal of Neurophysiology*, *101*, 773-88.

Corbetta, M. (1998). Frontoparietal cortical networks for directing attention and the eye to visual locations: identical, independent, or overlapping neural systems? *Proceedings of the National Academy of Sciences, U.S.A.*, *95*, 831-838.

Corbetta, M., & Shulman, G. L. (1998). Human cortical mechanisms of visual attention during orienting and search. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, *353*, 1353-1362.

Crone, E A, & Ridderinkhof, K. R. (2010). The developing brain: From theory to neuroimaging and back. *Developmental Cognitive Neuroscience*, *1*, 101-109.

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

Davis, E. T., & Palmer, J. (2004). Visual search and attention: An overview. *Spatial Vision*, *17*, 249-255.

den Ouden, H. E., Frith, U., Frith, C., & Blakemore, S. J. (2005). Thinking about intentions. *NeuroImage*, *28*, 787-796.

Dumontheil, I., & Blakemore, S-J. (in press). Social cognition and abstract thought in adolescence: The role of structural and functional development in rostral prefrontal cortex. *British Journal of Educational Psychology*.

Dumontheil, I., Apperly, I. A, & Blakemore, S. J. (2010). Online usage of theory of mind continues to develop in late adolescence. *Developmental Science*, *13*, 331-338.

Dumontheil, I., Hassan, B., Gilbert, S. J., & Blakemore, S.-J. (2010). Development of the selection and manipulation of self-generated thoughts in adolescence. *Journal of Neuroscience*, *30*, 7664-7671.

Dumontheil, I., Houlton, R., Christoff, K., & Blakemore, S.-J. (2010). Development of relational reasoning during adolescence. *Developmental Science*, *13*, F15-24.

Dumontheil, I., Küster, O., Apperly, I. A., & Blakemore, S.-J. (2010). Taking perspective into account in a communicative task. *NeuroImage*, *52*, 1574-1583.

Eisenberg, N., & Morris, A. S. (2004). Moral cognitions and prosocial responding in adolescence. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (pp. 155-188). Hoboken, NJ: Wiley.

Flavell, J., Abrahams Everett, B., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for Level 1-Level 2 distinction. *Developmental Psychology*, *17*, 99-103.

Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). "Who" is saying "what"? Brain-based decoding of human voice and speech. *Science, 322*, 970-973.

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C., Frackowiak, R. S., & Turner, R. (1995). Analysis of fMRI time-series revisited. *NeuroImage*, *2*, 45-53.

Frith, C. D., & Frith, U. (2007). Social cognition in humans. *Current Biology*, *17*, R724-32.

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

Fuster, J. M. (2000). Executive frontal functions. *Experimental Brain Research*, *133*, 66-70.

Gallagher, H. L., & Frith, Christopher D. (2003). Functional imaging of "theory of mind." *Trends in Cognitive Sciences*, *7*, 77-83.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., Paus, T., et al. (1999). Brain development during childhood and adolescence: A longitudinal MRI study. *Nature Neuroscience*, *2*, 861-863.

Giesbrecht, B., Woldorff, M. G., Song, A. W., & Mangun, G. R. (2003). Neural mechanisms of top-down control during spatial and feature attention. *NeuroImage*, *19*, 496-512.

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

Gunther Moor, B., Op de Macks, Z. A., Güroglu, B., Rombouts, S. A. R. B., Van der Molen, M. W., & Crone, E. A. (in press). Neurodevelopmental changes of reading the mind in the eyes. *Social Cognitive and Affective Neuroscience*.

Hahn, B., Ross, T. J., & Stein, E. A. (2006). Neuroanatomical dissociation between bottom-up and top-down processes of visuospatial selective attention. *NeuroImage*, *32*, 842-853.

Haxby, J., Hoffman, E., & Gobbini, M. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*, 223-233.

Humphreys, G. W., Allen, H. A., & Mavritsaki, E. (2009). Using biologically plausible neural models to specify the functional and neural mechanisms of visual search. *Progress in Brain Research*, *176*, 135-148.

Jenkins, A. C., Macrae, C. N., & Mitchell, J. P. (2008). Repetition suppression of ventromedial prefrontal activity during judgments of self and others. *Proceedings of the National Academy of Sciences, U.S.A., 105*, 4507-4512.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*, 32-38.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*, 25-41.

Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends in Cognitive Sciences*, *11*, 229-235.

Kovacs, A. M., Teglas, E., & Endress, A. D. (2010). The Social Sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*, 1830-1834.

Leslie, A. M. (2005). Developmental parallels in understanding minds and bodies. *Trends in Cognitive Sciences*, *9*, 459-462.

Luna, B., Padmanabhan, A., & O'Hearn, K. (2010). What has fMRI told us about the development of cognitive control through adolescence? *Brain and Cognition*, *72*, 101-113.

Mar, R. A. (2011). The neural bases of social cognition and story comprehension. *Annual Review of Psychology*, *62*, 103-134.

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron*, *50*, 655-663.

Moll, H., & Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *British Journal of Developmental Psychology2*, *24*, 603-613.

Moriguchi, Y., Ohnishi, T., Mori, T., Matsuda, H., & Komaki, G. (2007). Changes of brain activity in the neural substrates for theory of mind during childhood and adolescence. *Psychiatry and Clinical Neurosciences*, *61*, 355-363.

Noordzij, M. L., Newman-Norlund, S. E., de Ruiter, J. P., Hagoort, P., Levinson, S. C., & Toni, I. (2010). Neural correlates of intentional communication. *Frontiers in Neuroscience*, *4*, 188.

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.

Pfeifer, J. H., Masten, C. L., Borofsky, L. A., Dapretto, M., Fuligni, A. J., & Lieberman, M. D. (2009). Neural correlates of direct and reflected self-appraisals in adolescents and adults: When social perspective-taking informs self-perception. *Child Development*, *80*, 1016-1038.

Qureshi, A. W., Apperly, Ian A, & Samson, D. (2010). Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: evidence from a dual-task study of adults. *Cognition*, *117*, 230-236.

Samson, D., Apperly, I. A, Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology. Human Perception and Performance*, *36*, 1255-1266.

Samson, D., Apperly, I. A., Kathirgamanathan, U., & Humphreys, G. W. (2005). Seeing it my way: a case of a selective deficit in inhibiting self-perspective. *Brain*, *128*, 1102-1111.

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology*, *55*, 87-124.

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, *16*, 235-239.

Saxe, R. R., Whitfield-Gabrieli, S., Scholz, J., & Pelphrey, K. A. (2009). Brain regions for perceiving and reasoning about other people in school-aged children. *Child Development*, *80*, 1197-209.

Saxe, R., Schulz, L. E., & Jiang, Y. V. (2006). Reading minds versus following rules: dissociating theory of mind and executive control in the brain. *Social Neuroscience*, *1*, 284-298.

Saxe, R., & Powell, L. J. (2006). It’s the thought that counts: specific brain regions for one component of theory of mind. *Psychological Science*, *17*, 692-699.

Sebastian, C. L., Fontaine, N. M. G., Bird, G., Blakemore, S.-J., De Brito, S. A, McCrory, E. J. P., & Viding, E. (in press). Neural processing associated with cognitive and affective Theory of mind in adolescents and adults. *Social Cognitive and Affective Neuroscience*.

Shaw, P., Kabani, N. J., Lerch, J. P., Eckstrand, K., Lenroot, R., Gogtay, N., Greenstein, D., et al. (2008). Neurodevelopmental trajectories of the human cerebral cortex. *Journal of Neuroscience*, *28*, 3586-3594.

Sowell, E. R., Thompson, P. M., Holmes, C. J., Jernigan, T. L., & Toga, A. W. (1999). In vivo evidence for post-adolescent brain maturation in frontal and striatal regions. *Nature Neuroscience*, *2*, 859-861.

Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind and Language*, *17*, 3-23.

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience*, *10*, 640-656.

Tamir, D. I., & Mitchell, J. P. (2010). Neural correlates of anchoring-and-adjustment during mentalizing. *Proceedings of the National Academy of Sciences, U.S.A.*, *107*, 10827-10832.

Taylor, J. C., Wiggett, A. J., & Downing, P. E. (2007). Functional MRI analysis of body and body part representations in the extrastriate and fusiform body areas. *Journal of Neurophysiology*, *98*, 1626-1633.

Tong, F. (2003). Primary visual cortex and visual awareness. *Nature Reviews Neuroscience*, *4*, 219-229.

Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Human Brain Mapping*, *30*, 829-858.

Van Overwalle, F. (2011). A dissociation between social mentalizing and general reasoning. *NeuroImage*, *54*, 1589-1599.

Wang, A. T., Lee, S. S., Sigman, M., & Dapretto, M. (2006). Developmental changes in the neural basis of interpreting communicative intent. *Social Cognitive and Affective Neuroscience*, *1*, 107-121.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI)*. Harcourt Assessment, San Antonio, Texas.

7.4. Chapter 4

Aichhorn, M., Perner, J., Kronbichler, M., Staffen, W., Ladurner, G., 2006. Do visual perspective tasks need theory of mind? *NeuroImage* 30, 1059–1068.

Amodio, D.M., Frith, C.D., 2006. Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* 7, 268–277. doi:10.1038/nrn1884

Apperly, I.A., 2011. Mindreaders: The cognitive basis of "Theory of Mind". *Psychology Press*, Hove, UK.

Apperly, I.A., Carroll, D.J., Samson, D., Humphreys, G.W., Qureshi, A., Moffitt, G., 2010. Why are there limits on theory of mind use? Evidence from adults' ability to follow instructions from an ignorant speaker. *Q J Exp Psychol* (Colchester) 63, 1201–1217. doi:10.1080/17470210903281582

Beck, D.M., Kastner, S., 2009. Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vision research* 49, 1154–65. doi:10.1016/j.visres.2008.07.012

Bland, J.M., Altman, D.G., 1996a. Statistics Notes: Transforming data. *BMJ* 312, 770–770. doi:10.1136/bmj.312.7033.770

Bland, J.M., Altman, D.G., 1996b. Statistics notes: Transformations, means, and confidence intervals. *BMJ* 312, 1079–1079. doi:10.1136/bmj.312.7038.1079

Brothers, L., 1990. The social brain: A project for integrating primate behavior and neurophysiology in a new domain. *Concepts in Neuroscience* 1, 27–61.

Brown, E.C., Brüne, M., 2012. The role of prediction in social neuroscience. *Front Hum Neurosci* 6, 147. doi:10.3389/fnhum.2012.00147

Burgess, P.W., Gilbert, S.J., Dumontheil, I., 2007. Function and localization within rostral prefrontal cortex (area 10). *Philos Trans R Soc Lond B Biol Sci* 362, 887–899. doi:10.1098/rstb.2007.2095

Buschman, T.J., Miller, E.K., 2007. Top-Down Versus Bottom-Up Control of Attention in the Prefrontal and Posterior Parietal Cortices. *Science* 315, 1860 –1862. doi:10.1126/science.1138071

Clark, 2012. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*. Sec. 3, 1–86.

Daunizeau, J., Stephan, K.E., Friston, K., 2012. Stochastic dynamic causal modelling of fMRI data: Should we care about neural noise? *Neuroimage* 62, 464–481. doi:10.1016/j.neuroimage.2012.04.061

David, N., Bewernick, B.H., Cohen, M.X., Newen, A., Lux, S., Fink, G.R., Shah, N.J., Vogeley, K., 2006. Neural representations of self versus other: visual-spatial perspective taking and agency in a virtual ball-tossing game. *J Cogn Neurosci* 18, 898–910. doi:10.1162/jocn.2006.18.6.898

Dumontheil, I., Hillebrandt, H., Apperly, I.A., Blakemore, S.-J., 2012. Developmental Differences in the Control of Action Selection by Social Information. *Journal of Cognitive Neuroscience* 1–16. doi:10.1162/jocn_a_00268

Dumontheil, I., Küster, O., Apperly, I.A., Blakemore, S.-J., 2010. Taking perspective into account in a communicative task. *NeuroImage* 52, 1574–1583. doi:16/j.neuroimage.2010.05.056

Duncan, J., 2010. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends Cogn. Sci.* (Regul. Ed.) 14, 172–179. doi:10.1016/j.tics.2010.01.004

Friston, K., 2005. A theory of cortical responses. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 360, 815–836. doi:10.1098/rstb.2005.1622

Friston, K., 2010. The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi:10.1038/nrn2787

Friston, K., Daunizeau, J., Stephan, K.E., 2013. Model selection and gobbledygook: Response to Lohmann et al. *NeuroImage* 75, 275–278. doi:10.1016/j.neuroimage.2011.11.064

Friston, K., Harrison, L., Penny, W., 2003. Dynamic causal modelling. *Neuroimage* 19, 1273–1302.

Friston, K., Li, B., Daunizeau, J., Stephan, K.E., 2011. Network discovery with DCM. *Neuroimage* 56, 1202–1221. doi:10.1016/j.neuroimage.2010.12.039

Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., Penny, W., 2007. Variational free energy and the Laplace approximation. *Neuroimage* 34, 220–234. doi:10.1016/j.neuroimage.2006.08.035

Friston, K., Penny, W., 2011. Post hoc Bayesian model selection. *Neuroimage* 56, 2089–2099. doi:10.1016/j.neuroimage.2011.03.062

Frith, C.D., Frith, U., 2006. The Neural Basis of Mentalizing. *Neuron* 50, 531–534. doi:10.1016/j.neuron.2006.05.001

Frith, C.D., Frith, U., 2007. Social cognition in humans. *Curr. Biol* 17, R724–732. doi:10.1016/j.cub.2007.05.068

Frith, C.D., Frith, U., 2012. Mechanisms of Social Cognition. *Annual Review of Psychology* 63, 287–313. doi:10.1146/annurev-psych-120710-100449

Fuster, J., 2008. The Prefrontal Cortex, Fourth Edition, 4th ed. *Academic Press.*

Fuster, J.M., 2000. Executive frontal functions. *Exp. Brain Res.* 133, 66–70. doi:10.1007/s002210000401

Goulden, N., Elliott, R., Suckling, J., Williams, S.R., Deakin, J.F.W., McKie, S., 2012. Sample size estimation for comparing parameters using dynamic causal modeling. *Brain Connect* 2, 80–90. doi:10.1089/brain.2011.0057

Hahn, B., Ross, T.J., Stein, E.A., 2006. Neuroanatomical dissociation between bottom-up and top-down processes of visuospatial selective attention. *Neuroimage* 32, 842–853. doi:10.1016/j.neuroimage.2006.04.177

Hein, G., Knight, R.T., 2011. Superior Temporal Sulcus—It's My Area: Or Is It? *Journal of Cognitive Neuroscience* 20, 2125–2136. doi:i: 10.1162/jocn.2008.20148</p>

Ji, G., Neugebauer, V., 2012. Modulation of medial prefrontal cortical activity using in vivo recordings and optogenetics. *Mol Brain* 5, 36. doi:10.1186/1756-6606-5-36

Kasess, C.H., Stephan, K.E., Weissenbacher, A., Pezawas, L., Moser, E., Windischberger, C., 2010. Multi-subject analyses with dynamic causal modeling. *Neuroimage* 49, 3065–3074. doi:10.1016/j.neuroimage.2009.11.037

Keysar, B., Barr, D.J., Balin, J.A., Brauner, J.S., 2000. Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science* 11, 32–38.

Keysar, B., Lin, S., Barr, D.J., 2003. Limits on theory of mind use in adults. *Cognition* 89, 25–41.

Koechlin, E., Summerfield, C., 2007. An information theoretical approach to prefrontal executive function. *Trends in cognitive sciences* 11, 229–235.

Kret, M.E., Pichon, S., Grèzes, J., de Gelder, B., 2011. Similarities and differences in perceiving threat from dynamic faces and bodies. An fMRI study. *NeuroImage* 54, 1755–1762. doi:10.1016/j.neuroimage.2010.08.012

Levy, B.J., Wagner, A.D., 2011. Cognitive control and right ventrolateral prefrontal cortex: reflexive reorienting, motor inhibition, and action updating. *Annals of the New York Academy of Sciences* 1224, 40–62. doi:10.1111/j.1749-6632.2011.05958.x

Lin, S., Keysar, B., Epley, N., 2010. Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *J. Exp. Soc. Psychol*. 46, 551–556. doi:10.1016/j.jesp.2009.12.019

Lohmann, G., Erfurth, K., Müller, K., Turner, R., 2012. Critical comments on dynamic causal modelling. *NeuroImage* 59, 2322–2329. doi:10.1016/j.neuroimage.2011.09.025

Lui, L.L., Bourne, J.A., Rosa, M.G.P., 2006. Functional Response Properties of Neurons in the Dorsomedial Visual Area of New World Monkeys (Callithrix jacchus). *Cereb. Cortex* 16, 162–177. doi:10.1093/cercor/bhi094

Marreiros, A.C., Kiebel, S.J., Friston, K., 2008. Dynamic causal modelling for fMRI: a two-state model. *Neuroimage* 39, 269–278. doi:10.1016/j.neuroimage.2007.08.019

Meyer, M.L., Spunt, R.P., Berkman, E.T., Taylor, S.E., Lieberman, M.D., 2012. Evidence for social working memory from a parametric functional MRI study. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1883–1888. doi:10.1073/pnas.1121077109

Onitsuka, T., Shenton, M.E., Salisbury, D.F., Dickey, C.C., Kasai, K., Toner, S.K., Frumin, M., Kikinis, R., Jolesz, F.A., McCarley, R.W., 2004. Middle and Inferior Temporal Gyrus Gray Matter Volume Abnormalities in Chronic Schizophrenia: An MRI Study. *Am J Psychiatry* 161, 1603–1611. doi:10.1176/appi.ajp.161.9.1603

Penny, W., Stephan, K.E., Daunizeau, J., Rosa, M.J., Friston, K., Schofield, T.M., Leff, A.P., 2010. Comparing Families of Dynamic Causal Models. *PLoS Comput Biol* 6, e1000709. doi:10.1371/journal.pcbi.1000709

Penny, W., Stephan, K.E., Mechelli, A., Friston, K., 2004. Comparing dynamic causal models. *NeuroImage* 22, 1157–1172. doi:16/j.neuroimage.2004.03.026

Rosa, M.J., Friston, K., Penny, W., 2012. Post-hoc selection of dynamic causal models. *J. Neurosci. Methods* 208, 66–78. doi:10.1016/j.jneumeth.2012.04.013

Samson, D., Apperly, I.A., 2010. There is more to mind reading than having theory of mind concepts: new directions in theory of mind research. *Infant and Child Development* 19, 443–454. doi:10.1002/icd.678

Saxe, R., Schulz, L.E., Jiang, Y.V., 2006. Reading minds versus following rules: dissociating theory of mind and executive control in the brain. *Soc Neurosci* 1, 284–298. doi:10.1080/17470910601000446

Scholz, J., Triantafyllou, C., Whitfield-Gabrieli, S., Brown, E.N., Saxe, R., 2009. Distinct regions of right temporo-parietal junction are selective for theory of mind and exogenous attention. *PLoS ONE* 4, e4869. doi:10.1371/journal.pone.0004869

Sharp, D.J., Bonnelle, V., Boissezon, X.D., Beckmann, C.F., James, S.G., Patel, M.C., Mehta, M.A., 2010. Distinct frontal systems for response inhibition, attentional capture, and error processing. *PNAS*. doi:10.1073/pnas.1000175107

Stephan, K.E., Kasper, L., Harrison, L.M., Daunizeau, J., den Ouden, H.E.M., Breakspear, M., Friston, K., 2008. Nonlinear dynamic causal models for fMRI. *Neuroimage* 42, 649–662. doi:10.1016/j.neuroimage.2008.04.262

Stephan, K.E., Penny, W., Moran, R.J., den Ouden, H.E.M., Daunizeau, J., Friston, K., 2010. Ten simple rules for dynamic causal modeling. *Neuroimage* 49, 3099–3109. doi:10.1016/j.neuroimage.2009.11.015

Tamir, D.I., Mitchell, J.P., 2010. Neural correlates of anchoring-and-adjustment during mentalizing. *PNAS* 107, 10827–10832. doi:10.1073/pnas.1003242107

Van Overwalle, F., 2009. Social cognition and the brain: a meta-analysis. *Hum Brain Mapp* 30, 829–858. doi:10.1002/hbm.20547

Van Overwalle, F., 2011. A dissociation between social mentalizing and general reasoning. *Neuroimage* 54, 1589–1599. doi:10.1016/j.neuroimage.2010.09.043

Vogeley, K., May, M., Ritzl, A., Falkai, P., Zilles, K., Fink, G.R., 2004. Neural correlates of first-person perspective as one constituent of human self-consciousness. *J Cogn Neurosci* 16, 817–827. doi:10.1162/089892904970799

7.5. Chapter 5

Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., … Van Essen, D. C. (2013). Function in the Human Connectome: Task-fMRI and Individual Differences in Behavior. *NeuroImage*. doi:10.1016/j.neuroimage.2013.05.033

Barrett, J. L. (2000). Exploring the natural foundations of religion. *Trends in Cognitive Sciences*, *4*(1), 29–34. doi:10.1016/S1364-6613(99)01419-9

Bland, J. M., & Altman, D. G. (1996a). Statistics Notes: Transforming data. *BMJ*, *312*(7033), 770–770. doi:10.1136/bmj.312.7033.770

Bland, J. M., & Altman, D. G. (1996b). Statistics notes: Transformations, means, and confidence intervals. *BMJ*, *312*(7038), 1079–1079. doi:10.1136/bmj.312.7038.1079

Born, R. T., & Bradley, D. C. (2005). Structure and Function of Visual Area Mt. *Annual Review of Neuroscience*, *28*(1), 157–189. doi:10.1146/annurev.neuro.26.041002.131052

Brown, E. C., & Brüne, M. (2012). The role of prediction in social neuroscience. *Frontiers in human neuroscience*, *6*, 147. doi:10.3389/fnhum.2012.00147

Büchel, C., Holmes, A. P., Rees, G., & Friston, K. (1998). Characterizing Stimulus–Response Functions Using Nonlinear Regressors in Parametric fMRI Experiments. *NeuroImage*, *8*(2), 140–148. doi:10.1006/nimg.1998.0351

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

Carp, J. (2013). Better living through transparency: Improving the reproducibility of fMRI results through comprehensive methods reporting. *Cognitive, Affective, & Behavioral Neuroscience*, 1–7. doi:10.3758/s13415-013-0188-0

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.

Clark. (2012). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences.*, *Sec. 3*, 1–86.

Daunizeau, J., Stephan, K. E., & Friston, K. (2012). Stochastic dynamic causal modelling of fMRI data: Should we care about neural noise? *NeuroImage*, *62*(1), 464–481. doi:10.1016/j.neuroimage.2012.04.061

David, S. P., Ware, J. J., Chu, I. M., Loftus, P. D., Fusar-Poli, P., Radua, J., … Ioannidis, J. P. A. (2013). Potential Reporting Bias in fMRI Studies of the Brain. *PLoS ONE*, *8*(7), e70104. doi:10.1371/journal.pone.0070104

Egner, T., Monti, J. M., & Summerfield, C. (2010). Expectation and Surprise Determine Neural Population Responses in the Ventral Visual Stream. *The Journal of Neuroscience*, *30*(49), 16601–16608. doi:10.1523/JNEUROSCI.2770-10.2010

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, *25*(4), 1325–1335. doi:10.1016/j.neuroimage.2004.12.034

Feinberg, D. A., Moeller, S., Smith, S. M., Auerbach, E., Ramanna, S., Glasser, M. F., … Yacoub, E. (2010). Multiplexed Echo Planar Imaging for Sub-Second Whole Brain FMRI and Fast Diffusion Imaging. *PLoS ONE*, *5*(12), e15710. doi:10.1371/journal.pone.0015710

Feldman, H., & Friston, K. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*, 215. doi:10.3389/fnhum.2010.00215

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature reviews. Neuroscience*, *11*(2), 127–138. doi:10.1038/nrn2787

Friston, K. (2012). Ten ironic rules for non-statistical reviewers. *NeuroImage*, *61*(4), 1300–1310. doi:10.1016/j.neuroimage.2012.04.018

Friston, K., Adams, R. A., Perrinet, L., & Breakspear, M. (2012). Perceptions as hypotheses: saccades as experiments. *Frontiers in Perception Science*, *3*, 151. doi:10.3389/fpsyg.2012.00151

Friston, K., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*(4), 1273–1302.

Friston, K., Li, B., Daunizeau, J., & Stephan, K. E. (2011). Network discovery with DCM. *NeuroImage*, *56*(3), 1202–1221. doi:10.1016/j.neuroimage.2010.12.039

Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., & Penny, W. (2007). Variational free energy and the Laplace approximation. *NeuroImage*, *34*(1), 220–234. doi:10.1016/j.neuroimage.2006.08.035

Friston, K., & Penny, W. (2011). Post hoc Bayesian model selection. *NeuroImage*, *56*(4), 2089–2099. doi:10.1016/j.neuroimage.2011.03.062

Gao, T., Scholl, B. J., & McCarthy, G. (2012). Dissociating the detection of intentionality from animacy in the right posterior superior temporal sulcus. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *32*(41), 14276–14280. doi:10.1523/JNEUROSCI.0562-12.2012

Gilaie-Dotan, S., Kanai, R., Bahrami, B., Rees, G., & Saygin, A. P. (2013). Neuroanatomical correlates of biological motion detection. *Neuropsychologia*, *51*(3), 457–463. doi:10.1016/j.neuropsychologia.2012.11.027

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., … for the WU-Minn HCP Consortium. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*. doi:10.1016/j.neuroimage.2013.04.127

Goulden, N., Elliott, R., Suckling, J., Williams, S. R., Deakin, J. F. W., & McKie, S. (2012). Sample size estimation for comparing parameters using dynamic causal modeling. *Brain connectivity*, *2*(2), 80–90. doi:10.1089/brain.2011.0057

Grossman, E. D., Battelli, L., & Pascual-Leone, A. (2005). Repetitive TMS over posterior STS disrupts perception of biological motion. *Vision Research*, *45*(22), 2847–2853. doi:10.1016/j.visres.2005.05.027

Herrington, J. D., Nymberg, C., & Schultz, R. T. (2011). Biological motion task performance predicts superior temporal sulcus activity. *Brain and Cognition*, *77*(3), 372–381. doi:10.1016/j.bandc.2011.09.001

Hillebrandt, H., Dumontheil, I., Blakemore, S.-J., & Roiser, J. P. (In press). Dynamic Causal Modelling of effective connectivity during perspective taking in a communicative task. *NeuroImage*.

Kandel, E. R., Markram, H., Matthews, P. M., Yuste, R., & Koch, C. (2013). Neuroscience thinks big (and collaboratively). *Nature Reviews Neuroscience*, *14*(9), 659–664. doi:10.1038/nrn3578

Kasess, C. H., Stephan, K. E., Weissenbacher, A., Pezawas, L., Moser, E., & Windischberger, C. (2010). Multi-subject analyses with dynamic causal modeling. *NeuroImage*, *49*(4), 3065–3074. doi:10.1016/j.neuroimage.2009.11.037

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*(430), 773–795. doi:10.2307/2291091

Kiebel, S. J., Klöppel, S., Weiskopf, N., & Friston, K. (2007). Dynamic causal modeling: a generative model of slice timing in fMRI. *NeuroImage*, *34*(4), 1487–1496. doi:10.1016/j.neuroimage.2006.10.026

Klein, A. M., Zwickel, J., Prinz, W., & Frith, U. (2009). Animated triangles: an eye tracking investigation. *Quarterly journal of experimental psychology (2006)*, *62*(6), 1189–1197. doi:10.1080/17470210802384214

Koster-Hale, J., & Saxe, R. (2013). Theory of Mind: A Neural Prediction Problem. *Neuron*, *79*(5), 836–848. doi:10.1016/j.neuron.2013.08.020

Lee, S. M., Gao, T., & McCarthy, G. (2012). Attributing intentions to random motion engages the posterior superior temporal sulcus. *Social cognitive and affective neuroscience*. doi:10.1093/scan/nss110

Lewis, J. W., & Van Essen, D. C. (2000). Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *The Journal of Comparative Neurology*, *428*(1), 112–137. doi:10.1002/1096-9861(20001204)428:1<112::AID-CNE8>3.0.CO;2-9

Marreiros, A. C., Kiebel, S. J., & Friston, K. (2008). Dynamic causal modelling for fMRI: a two-state model. *NeuroImage*, *39*(1), 269–278. doi:10.1016/j.neuroimage.2007.08.019

Moeller, S., Yacoub, E., Olman, C. A., Auerbach, E., Strupp, J., Harel, N., & Uğurbil, K. (2010). Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magnetic resonance in medicine: official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, *63*(5), 1144–1153. doi:10.1002/mrm.22361

Nguyen, V. T., Breakspear, M., & Cunnington, R. (in press). Fusing concurrent EEG-fMRI with dynamic causal modeling: Application to effective connectivity during face perception. *NeuroImage*. doi:10.1016/j.neuroimage.2013.06.083

Pavlova, M. A. (2012). Biological Motion Processing as a Hallmark of Social Cognition. *Cerebral Cortex*, *22*(5), 981–995. doi:10.1093/cercor/bhr156

Penny, W., Stephan, K. E., Daunizeau, J., Rosa, M. J., Friston, K., Schofield, T. M., & Leff, A. P. (2010). Comparing Families of Dynamic Causal Models. *PLoS Computational Biology*, *6*(3), e1000709. doi:10.1371/journal.pcbi.1000709

Penny, W., Stephan, K. E., Mechelli, A., & Friston, K. (2004). Comparing dynamic causal models. *NeuroImage*, *22*(3), 1157–1172. doi:16/j.neuroimage.2004.03.026

Poldrack, R. A., Barch, D. M., Wager, T. D., Wagner, A. D., Devlin, J. T., & Milham, M. P. (2013). Toward open sharing of task-based fMRI data: the OpenfMRI project. *Frontiers in Neuroinformatics*, *7*, 12. doi:10.3389/fninf.2013.00012

Price, C. J., & Friston, K. (1997). Cognitive Conjunction: A New Approach to Brain Activation Experiments. *NeuroImage*, *5*(4), 261–270. doi:10.1006/nimg.1997.0269

Ramsey, R., & Hamilton, A. F. de C. (2010). Triangles have goals too: Understanding action representation in left aIPS. *Neuropsychologia*, *48*(9), 2773–2776. doi:10.1016/j.neuropsychologia.2010.04.028

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, *2*(1), 79–87. doi:10.1038/4580

Rosa, M. J., Friston, K., & Penny, W. (2012). Post-hoc selection of dynamic causal models. *Journal of neuroscience methods*, *208*(1), 66–78. doi:10.1016/j.jneumeth.2012.04.013

Stephan, K. E., Kasper, L., Harrison, L. M., Daunizeau, J., den Ouden, H. E. M., Breakspear, M., & Friston, K. (2008). Nonlinear dynamic causal models for fMRI. *NeuroImage*, *42*(2), 649–662. doi:10.1016/j.neuroimage.2008.04.262

Stephan, K. E., Penny, W., Moran, R. J., den Ouden, H. E. M., Daunizeau, J., & Friston, K. (2010). Ten simple rules for dynamic causal modeling. *NeuroImage*, *49*(4), 3099–3109. doi:10.1016/j.neuroimage.2009.11.015

Stephan, K. E., Tittgemeyer, M., Knösche, T. R., Moran, R. J., & Friston, K. (2009). Tractography-based priors for dynamic causal models. *NeuroImage*, *47*(4), 1628–1638. doi:10.1016/j.neuroimage.2009.05.096

Uğurbil, K., Xu, J., Auerbach, E. J., Moeller, S., Vu, A., Duarte-Carvajalino, J. M., … Yacoub, E. (2013). Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. *NeuroImage*. doi:10.1016/j.neuroimage.2013.05.012

Van Essen, D., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., & Ugurbil, K. (2013). The WU-Minn Human Connectome Project: An overview. *NeuroImage*, *80*, 62–79. doi:10.1016/j.neuroimage.2013.05.041

Van Essen, D., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T. E. J., Bucholz, R., … Yacoub, E. (2012). The Human Connectome Project: A data acquisition perspective. *NeuroImage*, *62*(4), 2222–2231. doi:10.1016/j.neuroimage.2012.02.018

7.6. Chapter 6

Adolphs, R. (2010). Conceptual Challenges and Directions for Social Neuroscience. *Neuron*, 65(6), 752–767. doi:10.1016/j.neuron.2010.03.006

Azevedo, F. A. C., Carvalho, L. R. B., Grinberg, L. T., Farfel, J. M., Ferretti, R. E. L., Leite, R. E. P., … Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *The Journal of Comparative Neurology*, 513(5), 532–541. doi:10.1002/cne.21974

Barrett, J. L. (2000). Exploring the natural foundations of religion. *Trends in Cognitive Sciences*, 4(1), 29–34. doi:10.1016/S1364-6613(99)01419-9

Blakemore, S.-J., & Mills, K. L. (2014). Is Adolescence a Sensitive Period for Sociocultural Processing? *Annual Review of Psychology*, 65(1), null. doi:10.1146/annurev-psych-010213-115202

Bonein, A., & Serra, D. (2009). Gender pairing bias in trustworthiness. Journal of Behavioral and Experimental Economics (formerly The Journal of Socio-Economics), 38(5), 779–789.

Bono, M. de, & Villu Maricq, A. (2005). Neuronal Substrates of Complex Behaviors in C. Elegans. *Annual Review of Neuroscience*, 28(1), 451–501. doi:10.1146/annurev.neuro.27.070203.144259

Borgwardt, S., Radua, J., Mechelli, A., & Fusar-Poli, P. (2012). Why are psychiatric imaging methods clinically unreliable? Conclusions and practical guidelines for authors, editors and reviewers. *Behavioral and Brain Functions*, 8(1), 46. doi:10.1186/1744-9081-8-46

Brodersen, K. H., Deserno, L., Schlagenhauf, F., Lin, Z., Penny, W. D., Buhmann, J. M., & Stephan, K. E. (2014). Dissecting psychiatric spectrum disorders by generative embedding. *NeuroImage: Clinical,* 4, 98–111. doi:10.1016/j.nicl.2013.11.002

Brodersen, K. H., Schofield, T. M., Leff, A. P., Ong, C. S., Lomakina, E. I., Buhmann, J. M., & Stephan, K. E. (2011). Generative embedding for model-based classification of fMRI data. *PLoS Computational Biology*, 7(6), e1002079. doi:10.1371/journal.pcbi.1002079

Buschman, T. J., & Miller, E. K. (2007). Top-Down Versus Bottom-Up Control of Attention in the Prefrontal and Posterior Parietal Cortices. *Science*, 315(5820), 1860 –1862. doi:10.1126/science.1138071

Cahill, L. (2006). Why sex matters for neuroscience. *Nature Reviews Neuroscience*, 7(6), 477–484. doi:10.1038/nrn1909

Chang, C., Cunningham, J. P., & Glover, G. H. (2009). Influence of heart rate on the BOLD signal: the cardiac response function. *NeuroImage*, 44(3), 857–869. doi:10.1016/j.neuroimage.2008.09.029

Clark. (2012). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*., Sec. 3, 1–86.

Cook, J. L., Press, C., Saygin, A. P., Kilner, J., & Blakemore, S. J. (under review). Dissociable neural processing of objective and subjective components of biological motion.

Cook, J., Saygin, A. P., Swain, R., & Blakemore, S.-J. (2009). Reduced sensitivity to minimum-jerk biological motion in autism spectrum conditions. *Neuropsychologia*, 47(14), 3275–3278. doi:10.1016/j.neuropsychologia.2009.07.010

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, 25(4), 1325–1335. doi:10.1016/j.neuroimage.2004.12.034

Eisenberger, N. I. (2012). The pain of social disconnection: examining the shared neural underpinnings of physical and social pain. *Nature Reviews Neuroscience*, 13(6), 421–434. doi:10.1038/nrn3231

Ergun, S. J., García-Muñoz, T., & Rivas, F. (2012). Gender differences in economic experiments. *Revista Internacional de Sociología*, 70(1), 99–111.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews. Neuroscience*, 11(2), 127–138. doi:10.1038/nrn2787

Friston, K., Adams, R. A., Perrinet, L., & Breakspear, M. (2012). Perceptions as hypotheses: saccades as experiments. *Frontiers in Perception Science*, 3, 151. doi:10.3389/fpsyg.2012.00151

Friston, K., Moran, R., & Seth, A. K. (2013). Analysing connectivity with Granger causality and dynamic causal modelling. *Current Opinion in Neurobiology*, 23(2), 172–178. doi:10.1016/j.conb.2012.11.010

Friston, K., Thornton, C., & Clark, A. (2012). Free-energy minimization and the dark-room problem. *Frontiers in Psychology*, 3, 130. doi:10.3389/fpsyg.2012.00130

Frith, C. D., & Frith, U. (2012). Mechanisms of Social Cognition. *Annual Review of Psychology*, 63(1), 287–313. doi:10.1146/annurev-psych-120710-100449

Greene, J., & Cohen, J. (2004). For the law, neuroscience changes nothing and everything. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 359(1451), 1775–1785. doi:10.1098/rstb.2004.1546

Harris, J. J., Reynell, C., & Attwell, D. (2011). The physiology of developmental changes in BOLD functional imaging signals. *Developmental Cognitive Neuroscience*, 1(3), 199–216. doi:10.1016/j.dcn.2011.04.001

Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A Hierarchy of Time-Scales and the Brain. *PLoS Comput Biol*, 4(11), e1000209. doi:10.1371/journal.pcbi.1000209

Kilner, J. M., Friston, K., & Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. *Cognitive Processing*, 8(3), 159–166. doi:10.1007/s10339-007-0170-2

Klein, A. M., Zwickel, J., Prinz, W., & Frith, U. (2009). Animated triangles: an eye tracking investigation. *Quarterly Journal of Experimental Psychology* (2006), 62(6), 1189–1197. doi:10.1080/17470210802384214

Lee, M. H., Smyser, C. D., & Shimony, J. S. (2013). Resting-State fMRI: A Review of Methods and Clinical Applications. *American Journal of Neuroradiology*, 34(10), 1866–1872. doi:10.3174/ajnr.A3263

Lieberman, M. D., & Cunningham, W. A. (2009). Type I and Type II error concerns in fMRI research: re-balancing the scale. *Social Cognitive and Affective Neuroscience*, 4(4), 423–428. doi:10.1093/scan/nsp052

Maye, A., Hsieh, C., Sugihara, G., & Brembs, B. (2007). Order in Spontaneous Behavior. *PLoS ONE*, 2(5), e443. doi:10.1371/journal.pone.0000443

Molina, V., Papiol, S., Sanz, J., Rosa, A., Arias, B., Fatjó-Vilas, M., … Fañanás, L. (2011). Convergent evidence of the contribution of TP53 genetic variation (Pro72Arg) to metabolic activity and white matter volume in the frontal lobe in schizophrenia patients. *NeuroImage*, 56(1), 45–51. doi:10.1016/j.neuroimage.2011.01.076

Montague, P. R., Berns, G. S., Cohen, J. D., McClure, S. M., Pagnoni, G., Dhamala, M., … Fisher, R. E. (2002). Hyperscanning: Simultaneous fMRI during Linked Social Interactions. *NeuroImage*, 16(4), 1159–1164. doi:10.1006/nimg.2002.1150

Mumford, J. A. (2012). A power calculation guide for fMRI studies. *Social Cognitive and Affective Neuroscience*, 7(6), 738–742. doi:10.1093/scan/nss059

Munakata, Y. (2000). Challenges to the Violation-of-Expectation Paradigm: Throwing the Conceptual Baby Out With the Perceptual Processing Bathwater? *Infancy*, 1(4), 471–477. doi:10.1207/S15327078IN0104_7

Nowak, M., & Sigmund, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner’s Dilemma game. *Nature*, 364(6432), 56–58. doi:10.1038/364056a0

Pettersson-Yeo, W., Allen, P., Benetti, S., McGuire, P., & Mechelli, A. (2011). Dysconnectivity in schizophrenia: where are we now? *Neuroscience and Biobehavioral Reviews*, 35(5), 1110–1124. doi:10.1016/j.neubiorev.2010.11.004

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, 59(3), 2142–2154. doi:10.1016/j.neuroimage.2011.10.018

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, 489(7416), 427–430. doi:10.1038/nature11467

Scholl, & Tremoulet. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, 4(8), 299–309.

Setoh, P., Wu, D., Baillargeon, R., & Gelman, R. (2013). Young infants have biological expectations about animals. *Proceedings of the National Academy of Sciences*, 110(40), 15937–15942. doi:10.1073/pnas.1314075110

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366. doi:10.1177/0956797611417632

Stephan, K. E., Marshall, J. C., Penny, W. D., Friston, K. J., & Fink, G. R. (2007). Interhemispheric Integration of Visual Processing during Task-Driven Lateralization. *The Journal of Neuroscience*, 27(13), 3512–3522. doi:10.1523/JNEUROSCI.4766-06.2007

Tzieropoulos, H. (2013). The Trust Game in neuroscience: A short review. *Social Neuroscience*, 8(5), 407–416. doi:10.1080/17470919.2013.832375

Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Human Brain Mapping*, 30(3), 829–858. doi:10.1002/hbm.20547

White, J. G., Southgate, E., Thomson, J. N., & Brenner, S. (1986). The Structure of the Nervous System of the Nematode Caenorhabditis elegans. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 314(1165), 1–340. doi:10.1098/rstb.1986.0056

Woo, C.-W., Krishnan, A., & Wager, T. D. (2014). Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations. *NeuroImage*, 91, 412–419. doi:10.1016/j.neuroimage.2013.12.058

Yook, K., Harris, T. W., Bieri, T., Cabunoc, A., Chan, J., Chen, W. J., … Sternberg, P. W. (2012). WormBase 2012: more genomes, more data, new website. *Nucleic Acids Research*, 40(D1), D735–D741. doi:10.1093/nar/gkr954