Self-other control as a candidate neurocognitive mechanism of typical and atypical social cognition.

Sowden, Sophie Lauren

Awarding institution:
King's College London

Self-other control as a candidate neurocognitive mechanism of typical and atypical social cognition

Sophie Lauren Sowden

MRC Social, Genetic and Developmental Psychiatry Centre,

Institute of Psychiatry, Psychology and Neuroscience

King's College London

Submitted for the degree of Doctor of Philosophy in Psychology and Neuroscience

2018

Supervisors: Prof. Geoffrey Bird and Prof. Francesca Happé

Abstract

Despite ever-growing interest in the 'social brain' and the search for the neural underpinnings of social cognition, we are yet to fully understand the basic neurocognitive mechanisms underlying complex social behaviours and their development. One such candidate mechanism is the control of neural representations of the self and of other people (Brass, Ruby, & Spengler, 2009; Spengler, von Cramon, & Brass, 2009a). For example, when taking another's perspective or controlling the tendency to imitate another's actions, key milestones in social cognitive development, it is necessary to *control* or *switch between* one's own self-generated representations, and representations elicited by others. Moreover, if a common, underlying neurocognitive mechanism to explain performance across social domains is found, this may shed important light on our understanding of disorders accompanied by broad social impairments. This thesis will take three main foci in its investigation of self-other control as a candidate neurocognitive mechanism of typical and atypical social cognition. Firstly, the thesis will provide an examination of behavioural indexes of self-other control using experimental tasks. Secondly, a potential neural basis for such a mechanism will be investigated using neurostimulation. Thirdly, it will be investigated whether disorders of social cognition, in particular autism, may involve atypical modulation of self and other representations (Cook & Bird, 2012; Sowden & Shah, 2014) and thus may be characterised as 'disorders of self-other control' or top-down modulation of social behaviour. Using the findings from this thesis, discussion will be introduced about the potential for training such a mechanism in order to achieve across-domain amelioration of symptoms in individuals with such a profile of social impairments.

Table of Contents

Acknowledgments

I am grateful to the Medical Research Council for granting me a studentship and supporting the completion of this PhD. I am also hugely appreciative of all the individuals who participated in the studies in this thesis, without whom this would not have been possible.

Particular thanks go to my lab team and collaborators, with whom it has been the greatest pleasure to work. I could not have asked to work with a more friendly, supportive and constructive group of individuals. This includes Punit Shah for his motivation and support to me during my academic studies, as well as other collaborators including Michael Banissy, Gordon Wright, Rebecca Brewer and Jennifer Cook for their help with the development and interpretation of my work. I must also extend gratitude to Caroline Catmur. Firstly, for her belief in my abilities during my Undergraduate degree and providing me with the best possible platform from which to develop my passion for Psychology, but also for her unlimited and selfless support throughout my PhD. Thanks also go to Francesca Happé for her theoretical input in my work. Finally, the biggest thanks go to my brilliant supervisor Geoff Bird. His passion and knowledge in experimental psychology continue to amaze and inspire me and his guidance and support has extended well beyond the exceptional academic input he has had in my PhD work. It has been the biggest pleasure and a huge amount of fun to be Geoff's student and this I will definitely miss the most.

I owe so much to my wonderful parents, Kerry and Mike and my sister Emma, who have always believed in me and provided me with the drive and determination to work hard and achieve academically, but more importantly for their infinite love, encouragement and support in everything I do. They have always been a constant source of motivation for me. I am also hugely grateful to Bruno for helping me to believe in myself and for all his love and support.

Finally, I am enormously thankful to my friends. I cannot acknowledge everyone, but I would particularly like to thank all the friends I have made during my PhD at the SGDP. Completing this PhD would not have been possible, or even half as enjoyable, without the fun and tea-filled time spent with Bao, Laura, Toby, Luke, Vikki, Jessie, Jo, Jane, Tegan, Jenny, Hannah, and my office mate, best friend and all-round life support, Talar.

I dedicate this thesis to our dearly missed friend Baocong Xia.

Author Contributions

Chapter 2: Study design was conducted by myself and Geoff Bird. I created the stimuli (with help from Caroline Catmur), programmed the experimental paradigm and carried out the data collection for Experiment 1. Experiment 2 data collection was done by Svenja Koene. I conducted all data analysis for Experiments 1 and 2 and drafted the manuscript for publication, which was then edited by all co-authors.

Chapter 3: Study design was conducted by myself and GB. I created the stimuli and programmed the experimental paradigm. I collected and analysed the data. All authors were involved in the interpretation of the analyses and I drafted the manuscript for publication, which was then edited by all co-authors.

Chapter 5: Study design was conducted by all authors. Gordon Wright created the stimuli and programmed the experimental paradigm used in both Experiments 1 and 2. Brain stimulation expertise and equipment was provided by Michael Banissy. I completed data collection and data analysis. All authors were involved in interpretation of the analyses and I drafted the manuscript for publication, which was then commented on by all co-authors.

Chapter 6: Study design was conducted by myself and GB. Stimuli were created, and the experimental paradigm was programmed, by GW. I collected and analysed the data and data interpretation was carried out by myself, GB and CC. I drafted the chapter which was then commented on by GB and Francesca Happé.

Chapter 7: Study design was conducted by myself, Jennifer Cook, GB and CC. I programmed the experimental paradigm and carried out data collection with the assistance of Elisa Militaru, Sofia Koletsi and Eva Lymberopoulos. I completed all data analysis and all authors were involved in interpretation of these analyses. I drafted the chapter, which was then commented on by GB and FH.

About this thesis. This thesis is submitted incorporating published papers. Chapters 2, 3 and 5 are exact copies of published journal articles and are available on the relevant journal websites. Chapters 4, 6 and 7, along with the general introduction and discussion sections were written specifically for this thesis. For consistency, each chapter has its own reference list. All experiments included in this thesis were approved by local research ethics committees.

Table of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Table of Tables

Chapter 1

Chapter 2

Chapter 6

List of Abbreviations

The following abbreviations will be referred to throughout this thesis and are redefined at first use in each new chapter:

ADOS	Autism Diagnostic Observation Schedule
ANOVA	Analysis of variance
ASD	Autism Spectrum Disorder
AQ	Autism Spectrum Quotient
AQ-10	10-item Autism Spectrum Quotient
fMRI	Functional magnetic resonance imaging
LTD	Long-term depression
LTP	Long-term potentiation
MEP	Motor-evoked potential
MNS	Mirror neuron system
MO	Mid-occipital
mPFC	Medial prefrontal cortex
NMDA	N-methyl-D-aspartate
RT	Reaction time
TAS-20	Toronto Alexithymia Scale
TPJ	Temporoparietal junction
tDCS	Transcranial direct current stimulation
TMS	Transcranial magnetic stimulation
SD	Standard deviation
SEM	Standard error of the mean

Chapter 1. Introduction

Humans are uniquely social beings and crave the company of, and interaction with, their conspecifics (Adolphs, 2003). Social cognition refers to the cognitive processes involved in such social behaviour; how one is able to make sense of the social world, including attending to, perceiving, interpreting and responding to the actions, beliefs, desires, intentions and emotions of both the self and of others (Moskowitz, 2005). The proliferation of interest in this area of research has revealed the cognitive and neural underpinnings of key milestones in social cognitive development, as well as beginning to shed light on both the aetiology and consequences of atypical social cognition (Blakemore, Winston, & Frith, 2004). Identifying commonalities in the mechanisms recruited across various domains of social cognition is important here, and may provide an understanding not only of typical social cognitive function but also what happens when these mechanisms go awry. This thesis aims to investigate the possibility of a common, underlying neurocognitive mechanism for social cognition; how this may help to explain performance across social domains, as well as providing a framework for understanding the broad-reaching social impairments observed in developmental and clinical disorders such as autism and schizophrenia. A candidate process here is the ability to represent and switch between neural representations pertaining to both the self and to other people. Let us first consider some key areas of social cognition in which shared representations of the self and the other are fundamental.

1.1. Shared Representations

1.1.1. Imitation. A fundamental aspect of social interaction is our ability to mirror or imitate the actions of others, described as the 'social glue' (Lakin, Jefferis, Cheng, & Chartrand, 2003) required to create and maintain successful relationships with others. This has been highlighted in the field of social psychology, with classic observation studies

revealing a ‘chameleon effect’ where individuals unconsciously mimic the gestures and postures of those with whom they interact (Chartrand & Bargh, 1999; Chartrand, Cheng, & Jefferis, 2002).

Cognitive neuroscientists have sought a neural basis for such automatic imitation of others’ behaviour and proposed that it may rely on the human mirror neuron system (MNS). First discovered in monkeys (di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992; Rizzolatti & Craighero, 2004) and subsequently in humans (Molenberghs, Cunnington, & Mattingley, 2012), the MNS comprises regions in the premotor and parietal cortices of the brain which respond both during the performance and observation of the same action. Experimental studies demonstrating automatic imitation have been used as an index of the functioning of the MNS (Brass, Bekkering, Wohlschläger, & Prinz, 2000; Heyes, Bird, Johnson, & Haggard, 2005; Stürmer, Aschersleben, & Prinz, 2000). Thus, the MNS gives us one example of a shared representational system between ‘self’ and ‘other’, as it maps representations of others’ actions onto those of our own (Iacoboni et al., 1999).

1.1.2. Theory of mind, perspective-taking and empathy. Similarly, another key domain of social cognition for which simultaneous representation of ‘self’ and ‘other’ is particularly important is that of theory of mind. This concerns the ability to attribute mental states to both the self and others, recognising that two individuals may simultaneously hold different beliefs, desires and intentions in any given social situation (Premack & Woodruff, 1978). Moreover, throughout social interactions, humans constantly represent their own visual perspectives on the world, but also those of the individuals with whom they interact. Like in the motor domain, representations of mental states and perspectives of the self and other have been shown to recruit overlapping brain regions, found to centre around the medial prefrontal cortex (mPFC; Anderson, Bechara, Damasio, Tranel, & Damasio, 1999; Ferstl & von Cramon, 2002; Lombardo et al., 2010a; Saxe, Moran, Scholz, & Gabrieli, 2006).

Finally, the ability to empathise with our conspecifics is yet another ability necessary for mature and successful social cognition (Harris, de Rosnay, & Pons, 2005). Despite the complex and multi-faceted nature of emotion understanding, basic empathy requires an individual to comprehend the affective states of both the self and of those with whom they interact. There is a body of work concerning common neural networks recruited during directly experienced pain and empathy for others' pain (Jackson, Brunet, Meltzoff, & Decety, 2006; Lamm, Decety, & Singer, 2011; Obhi & Cross, 2016), as well as evidence from functional magnetic resonance imaging (fMRI; Hennenlotter et al., 2005; Singer et al., 2004; Wicker et al., 2003) and lesion studies (Calder, Keane, Manes, Antoun, & Young, 2000; Hutchinson, Davis, Lozano, Tasker, & Dostrovsky, 1999) that key regions involved in emotion processing for the self show overlap with processing emotions of others. These include motor and somatosensory areas of the brain but also the anterior insula and anterior cingulate cortex.

Although the factor structure and relation between different domains is only beginning to be revealed (Happé, Cook, & Bird, 2017), we see commonalities here in each of these domains as individuals simultaneously hold neural representations pertaining to both the self and to other people (Obhi & Cross, 2016). If we have such shared representational systems which recruit overlapping brain regions (Decety & Sommerville, 2003; Keysers & Gazzola, 2009), why is it that we do not perform every action we observe or confuse the pain or emotional state experienced by another individual as belonging to ourselves? As well as a shared representational system, humans must also possess a means of controlling self and other representations.

1.2. A Candidate Mechanism

Social interaction in a variety of domains appears dependent on a shared representational system. However, success in a social situation sometimes requires an individual to distance themselves from other people and in other instances requires one to engage more with their representations of others. For example, when taking another's perspective, engaging a successful theory of mind, or empathising with others, it is important to put aside or inhibit one's own perspective, mental or affective state and enhance that of the interacting other. This is evident in the moral judgement task used by Young and colleagues (Young, Camprodon, Hauser, Pascual-Leone, & Saxe 2010) which requires the participants to assess the morality of a character in putting sugar labelled as 'toxic' into her friend's coffee. The participant knows that the powder is in fact sugar, but the character believes it is toxic. Therefore, in successfully assessing the morality of the character's action, the observer must enhance the representation of the character's mental state and inhibit that of their own. Conversely, when partaking in social interaction, to control the tendency to imitate others' actions and instead generate our own independent actions, we must inhibit the motor representation we hold for the interacting other and activate the motor representation for our own intended action (Brass, Derrfuss, & von Cramon, 2005; Brass et al., 2009). These examples are discrepant in their requirement for self- versus other-representation enhancement – perspective taking, theory of mind and empathy require the other to be enhanced and the self inhibited, while the control of imitation requires the self to be enhanced and the other inhibited. It appears therefore, that it may be crucial to have the ability to *control* or *switch between* neural representations attributed to the self and to other people (Brass & Heyes, 2005; Cook, 2014; Decety & Sommerville, 2003; Spengler et al., 2009a).

As we see the requirement for control across social domains, it is important to note that this mechanism does not modulate specifically at the motor or sensory level. Herein

referred to as ‘self-other control’, such a mechanism explains how, depending on task relevance and task demands we are able to enhance one representation whilst inhibiting the other, via a domain-general attentional modulation of competing self- and other-related representations (Cook, 2014; Sowden & Shah, 2014). Accordingly, despite the very different high-level cognitive processes involved in a wide range of social cognitive abilities, it appears that many may rely on this common mechanism of self-other control.

1.2.1. Measuring self-other control. A task now readily used as a behavioural index of self-other control is that of the control of imitation (Figure 1; Brass, Bekkering, & Prinz, 2001; Brass et al., 2005, 2009; Catmur & Heyes, 2011; Santiesteban, Banissy, Catmur, & Bird, 2012a, 2015a; Santiesteban, Bird, Tew, Cioffi, & Banissy, 2015b; Santiesteban et al., 2012b; Sowden & Catmur, 2015; Spengler et al., 2009a). The task concerns inhibiting imitative response tendencies (thus is often called the ‘imitation-inhibition’ task) and requires subjects to produce body movements (often finger abductions along the vertical or horizontal plane) in response to a non-social, visual cue (such as a number or coloured square). Task irrelevant hand stimuli are simultaneously observed and perform movements which are either imitatively compatible or incompatible with those to be performed by the subject. Response time differences between compatible and incompatible trials are considered an index of an individual’s ability to control the competing representations of self and other; exciting the self-representation, whilst inhibiting the other-representation. Chapters 2 (Sowden, Koehne, Catmur, Dziobek, & Bird, 2016a) and 3 (Sowden, Brewer, Catmur, & Bird, 2016b) of the current thesis utilise this task as an index of self-other control.

Additionally, Obhi and Hogeveen (2013) have proposed a complimentary task which allows the inspection of performance under the opposite control requirements; *inhibiting* the self-representation whilst exciting the other-representation. During 50% of trials where the (previously) task-irrelevant hand stimulus is observed to perform a movement, individuals

must inhibit the cued action and instead perform that of the task-irrelevant hand. In combination, these complimentary tasks provide a neat index of control; the ability to supress not only representations of the other but also of the self.

A question arising from this, however, is to what extent such a mechanism of self-other control extends across domains of social cognition. Despite the very different higher-level cognitive processes involved in a wide range of social cognitive abilities, a series of behavioural findings in neurotypical adults support the existence of a common low-level mechanism of self-other control. Performance in one social domain such as the control of imitation correlates highly with performance in other social domains requiring self-other control. These include perspective-taking, theory of mind and empathy (Spengler, Bird, & Brass, 2010a), and remain even when controlling for more general executive functioning processes (e.g., Spengler, von Cramon, & Brass, 2010b).

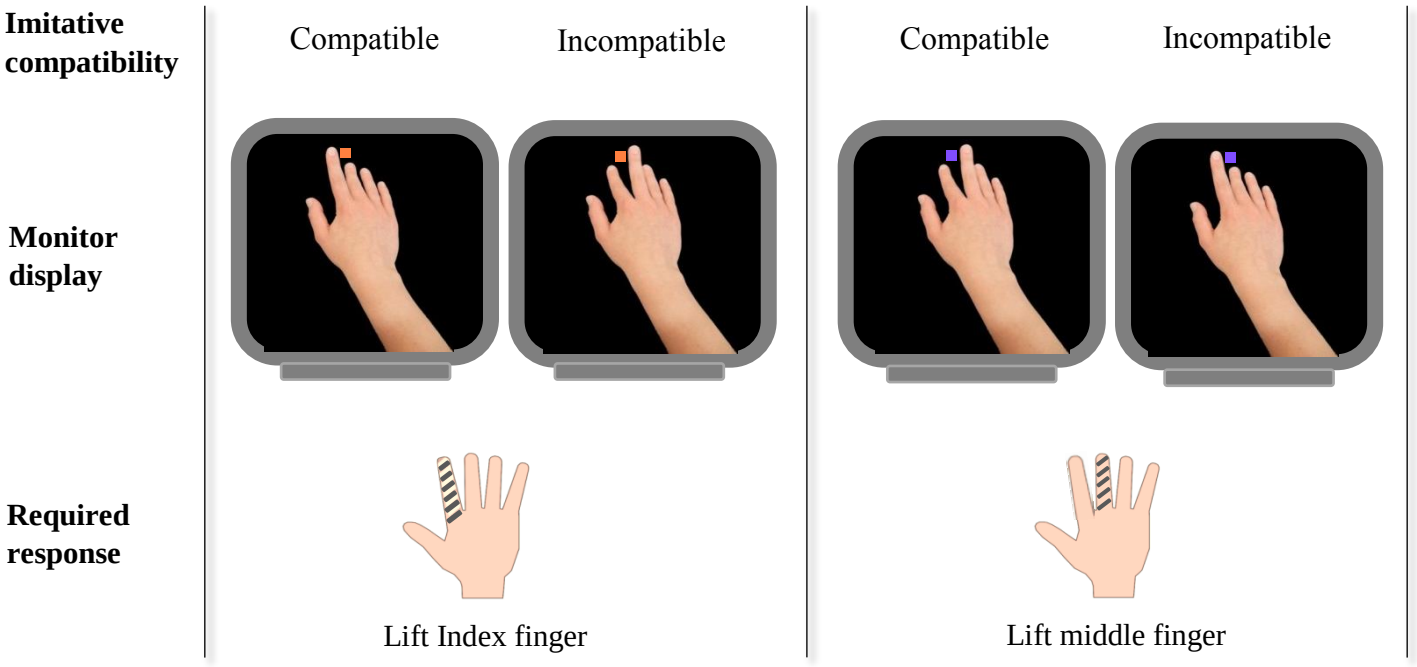

Figure 1. Examples of task-irrelevant hand stimuli and task-relevant cues presented during a task to measure the control of imitation or the imitation-inhibition task. Imitative compatibility of the task-irrelevant stimuli with the required finger lift response is also indicated for trials in which an orange square indicates lift index and a purple square indicates lift middle. Adapted from Sowden and Shah (2014).

The perspective-taking task is worthy of outlining in more detail here as it has also been used as an index of the integrity of one's self-other control (Santiesteban et al., 2012a, 2012b, 2015a, 2015b). In the Director Task (Figure 2; Keysar, Barr, Balin, & Brauner, 2000) the participant views a block of shelves. Standing on the other side of the shelves is a person (the 'director') and a number of the shelves are occluded from the director's viewpoint, whilst visible to the participant. Therefore, when instructed by the director to move objects around the shelves (up, down, left or right), to demonstrate good visual perspective-taking and thus self-other control, the participant must inhibit the self visual perspective and enhance that of the director (other), moving only the objects which are visible to the director and in the instructed direction with respect to the director's viewpoint. Crucially, Santiesteban and colleagues, in two separate samples of healthy controls (Santiesteban et al., 2012b, 2015b), found that performance on the imitation-inhibition and visual perspective-taking tasks was correlated.

The link between performance on different tasks requiring self-other control is not merely correlational; just 40 minutes of training to inhibit imitation produced an enhancement of perspective-taking ability (Santiesteban et al., 2012b), as well as increases in empathic corticospinal responses and self-reported empathy (de Guzman, Bird, Banissy, & Catmur, 2016). Participants were trained to inhibit the other- and enhance the self-representation, and this led to enhancements in social domains requiring inhibition of the self- and enhancement of the other-representation. Moreover, priming prosocial attitudes enhances automatic imitation but not a non-imitative control process (Cook & Bird, 2011; Leighton, Bird, Orsini, & Heyes, 2010), and engaging in more social interaction (Hogeveen & Obhi, 2012) or merely being imitated (Hogeveen, Chartrand, & Obhi, 2015) leads to increased neural representation of others. These examples all support the enhancement of a common process involved in social functioning. It also appears specific to social cognition as such modulation was not

observed on equivalent, non-social control tasks (Cook & Bird, 2011; Leighton et al., 2010; Santiesteban et al., 2012b).

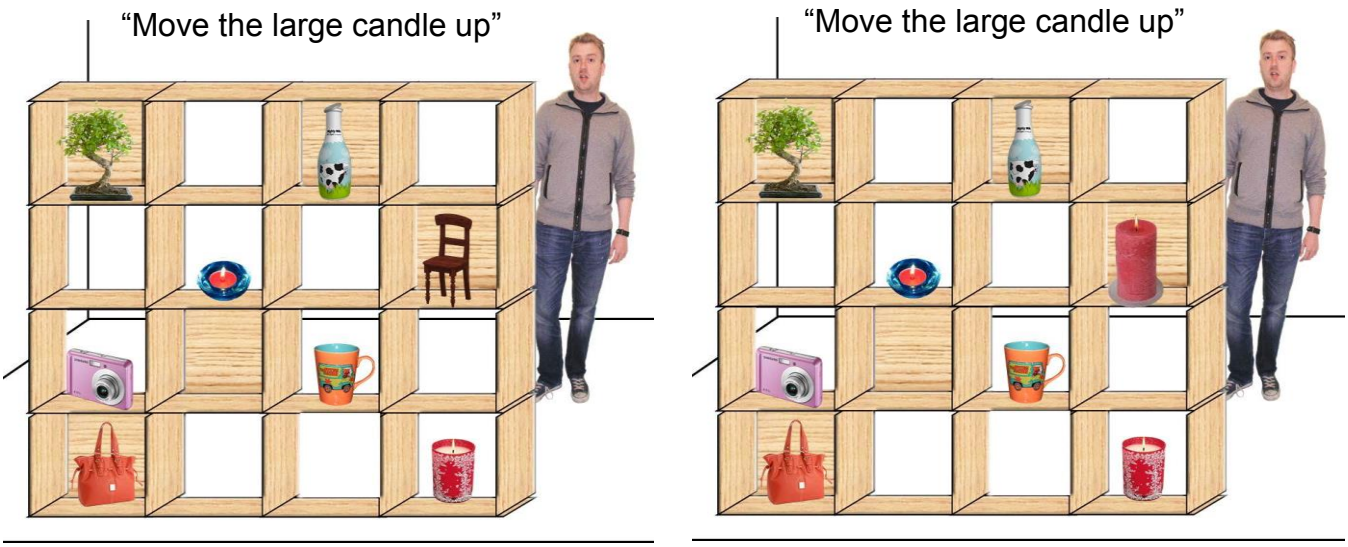

Figure 2. Example stimuli from the 'Director Task'; visual perspective-taking task. On the left is a control, no conflict trial where there is no conflict between the visual perspective of the self and the other in relation to the item instructed to be moved. On the right is an experimental trial where there is a conflict between the visual perspective of the self and the other in relation to the object instructed to be moved. In this situation, the participant must inhibit the representation they hold for their own visual perspective and enhance that of the director in order to move the correct object. Measures of performance relate to response times and accuracy to complete the instructed movement on each trial.

Thus, we have seen two robust indexes of self-other control, whereby the task elicits the online representation of self and other and which require the control or modulation of these representations for successful performance. Further social tasks which elicit similar cognitive demands could be used to demonstrate that self-other control is key across social domains. Chapter 5 (Sowden, Wright, Banissy, Catmur, & Bird, 2015), Experiment 1 of this thesis concerns this, extending self-other control to other, more high-level domains of social cognition and this chapter describes a novel task designed to elicit online self-other

interference effects in the context of opinions and lie detection.[1] The same task is then used to probe at the neural basis of this mechanism (Chapter 5, Experiment 2) and functioning of self-other control in individuals with Autism Spectrum Disorder (ASD; Chapter 6).

1.2.2. A neural basis for self-other control: The temporoparietal junction. As well as the mPFC, the temporoparietal junction (TPJ), a brain region located at the intersection of the superior temporal sulcus, lateral occipital cortex and inferior parietal lobule (Figure 3; Mars et al., 2012), has attracted extensive research attention over the past decade, emerging as a key node within the 'social brain' network (Frith & Frith, 2010; see Table 1 for more details on studies). Meta-analytic investigations of the function of the TPJ, focusing mainly on the *right* TPJ, have revealed its important role in a wide range of social cognitive abilities, including judging agency, perspective-taking, theory of mind and empathy (Figure 4; Decety & Lamm, 2007; Decety & Sommerville, 2003; Sperduti, Delaveau, Fossati, & Nadel, 2011; van Overwalle, 2009). These overlapping neural correlates indeed support the existence of a common neurocognitive mechanism and accordingly, Brass and colleagues (Brass et al., 2001, 2005, 2009; Spengler et al., 2009a, 2009b) utilised fMRI to localise the mechanism of self-other control, as indexed by the control of imitation, to the right TPJ and mPFC. The authors suggest both areas to be implicated in overcoming interference from observed behaviour, but the right TPJ to be particularly important in determining whether the action is performed by the self or the other, with the mPFC primarily managing and enforcing our own motor representations and actions as a result.

[1] By 'high-level' we are referring to domains involving more than just automatic, unconscious processes, but which may involve more conscious reasoning and interpretation and perhaps higher working memory demands.

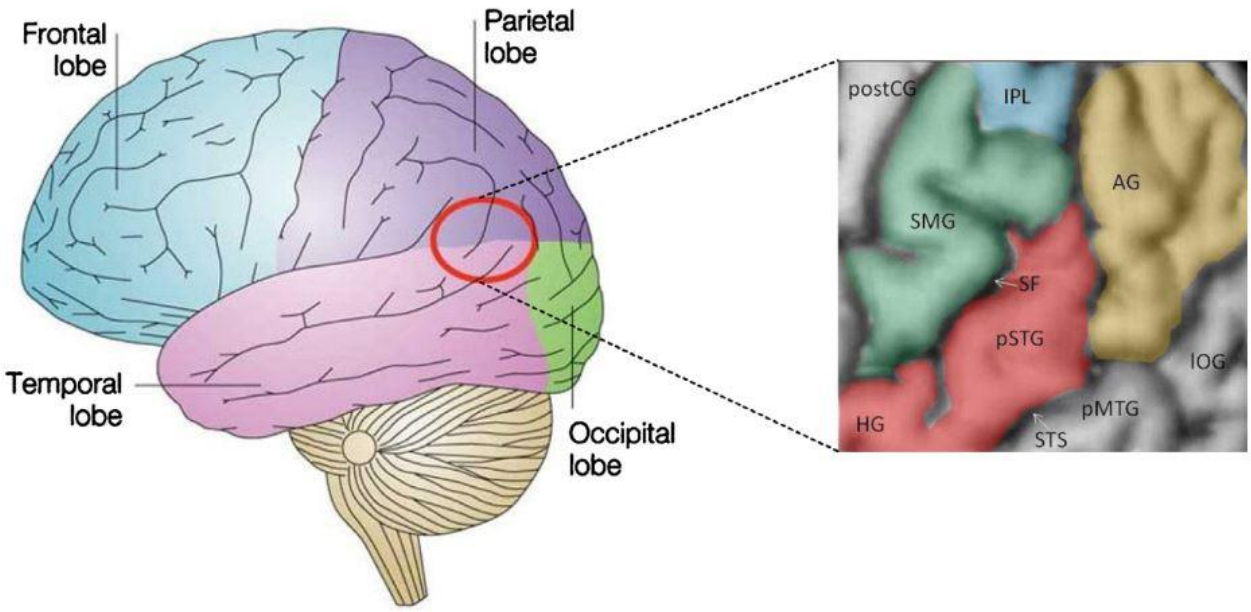

Figure 3. Visual depiction of the location of the temporoparietal junction, at the intersection of the superior temporal sulcus, lateral occipital cortex and inferior parietal lobule**.** Adapted from Tallal (2004).

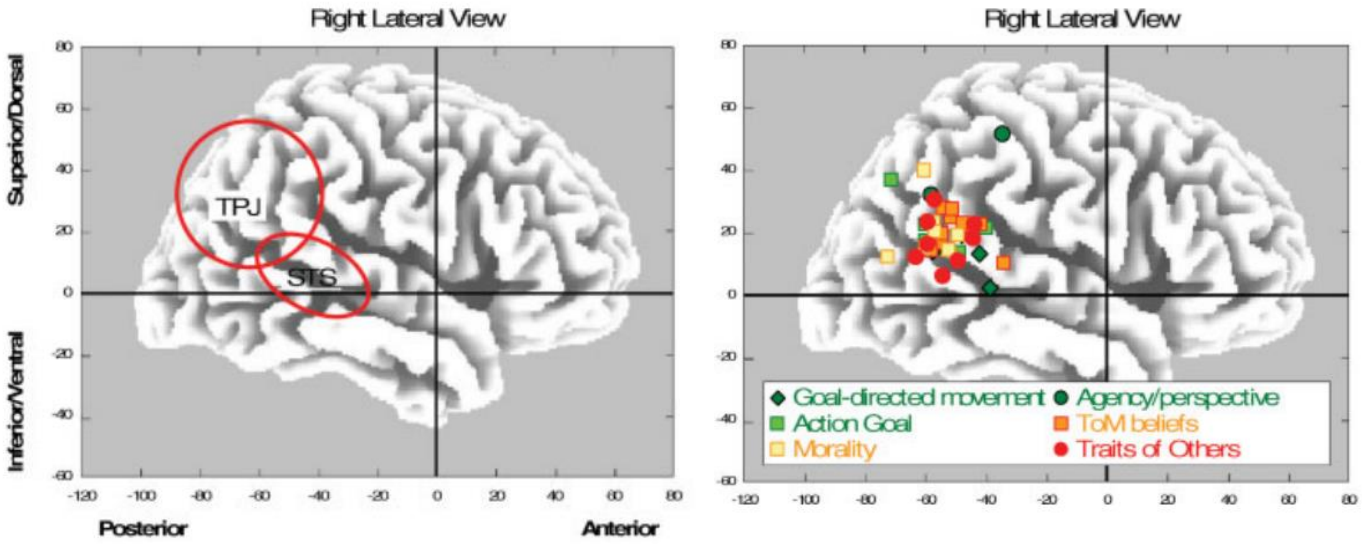

Figure 4. The anatomy of the human brain in x-y-z stereotactic atlas. Showing (left) the right temporoparietal junction (TPJ) and superior temporal sulcus (STS) and (right) the areas in the right TPJ recruited in studies investigating various social cognitive abilities. Adapted from van Overwalle (2009).

Table 1

Relevant Literature and Neuroscientific Methods Supporting the Role of the Right Temporoparietal junction in Various Social Functions.

Social function	Relevant supporting literature	Method
Empathy	Botvinick et al. (2005); Jackson et al. (2005, 2006); Lamm, Batson, & Decety (2007); Lawrence et al. (2006); Moriguchi et al. (2007); Singer et al. (2004); Voellm et al. (2006)	fMRI
	Decety & Chaminade (2003); Ruby & Decety (2004)	PET
Control of imitation	Brass et al. (2001, 2003, 2005, 2009); Spengler et al. (2009a, 2010a); Williams et al. (2006)	fMRI
	Chaminade, Meltzoff, & Decety (2002); Decety et al. (2002)	PET
	Hogeveen et al. (2015); Nobusako et al. (2017); Santiesteban et al. (2012a, 2015a)	tDCS
	Sowden & Catmur (2015)	TMS
Judging agency	Chaminade & Decety (2002); Farrer et al. (2003, 2004); Ruby & Decety (2001); Spence et al. (1997)	PET
	Farrer & Frith (2002); Kable & Chatterjee (2006); Leube et al. (2003); Ramnani & Miall (2004); Saxe et al. (2004); Schnell et al. (2007); Spengler et al. (2009b); Sperduti et al. (2011)	fMRI
Perspective-taking	David et al. (2006); Hynes, Baird, & Grafton (2006); Ruby & Decety (2001, 2003)	fMRI
	Nobusako et al. (2017); Santiesteban et al. (2012a)	tDCS
Theory of mind	Baron-Cohen et al. (1999a); Castelli et al. (2000); den Ouden et al. (2005); Gallagher & Frith (2003); Gallagher et al. (2000); Grèzes, Frith, & Passingham (2004); Moriguchi et al. (2006); Ohnishi et al. (2004); Perner et al. (2006); Rilling et al. (2004); Saxe (2009, 2010); Saxe & Kanwisher (2003); Saxe & Powell (2006); Saxe & Wexler (2005); Scholz et al. (2009); Spengler et al. (2009a); Vogeley et al. (2001); Walter et al. (2004); Young et al. (2007)	fMRI
	Brunet et al. (2000, 2003); Fletcher et al. (1995)	PET
	Bardi et al. (2017); Costa et al. (2008); Young et al. (2010)	TMS
	Mai et al. (2016)	tDCS
Self/other body representation	Cazzato et al. (2015); Tsakiris et al. (2008)	TMS
Lie detection	Sowden et al. (2015)	tDCS

Further, causal evidence for the role of the right TPJ in self-other control is derived from studies measuring the effects of magnetic or electrical stimulation of this area. Disruptive repetitive transcranial magnetic stimulation (TMS) of the right TPJ has been shown to impair performance in both the control of imitation (Sowden & Catmur, 2015) and theory of mind (Bardi, Six, & Brass, 2017; Costa, Torriero, Oliveri, & Caltagirone, 2008; Young et al., 2010). Anodal (excitatory) transcranial direct current stimulation (tDCS) enhanced imitative control and perspective-taking performance, whilst cathodal and sham stimulation to the same area, and active control stimulation to an alternative area, had no effect on performance in these tasks. This excitatory effect has now been replicated across 3 studies (Hogeveen et al., 2015; Santiesteban et al., 2012a, 2015a) and this stimulation has been shown to be effective when stimulating both *left* and *right* TPJ (Santiesteban et al., 2015a). Moreover, the same protocol of anodal tDCS to the right TPJ resulted in enhanced theory of mind performance in another study (Mai et al., 2016). The work of Santiesteban and colleagues is particularly noteworthy here, whereby the paradox of improved performance in these two tasks results from the opposing modulation of representations of self and other. Excitation of the TPJ promoted an enhanced representation of the self relative to the other in the imitation task and the other relative to the self in the perspective-taking task. As with the previous observation of behavioural training in one social domain leading to enhanced performance in another domain where there are opposing requirements of self versus other enhancement/inhibition, this pattern of results is best explained by the up-regulation of a mechanism which facilitates the *control* of self and other representations. Thus, it appears to be a dynamic mechanism capable of modulating behaviour based on the specific demands of a given social situation or experimental task.

Similarly, acquired temporoparietal lesions have been associated with rare disorders such as asomatognosia, characterised by the misidentification of part of one's own body as

belonging to another (Feinberg, Venneri, Simone, Fan, & Northoff, 2010) and anosognosia, characterised by a denial or unawareness of a paralysed limb (Ramachandran & Blakeslee, 1998). The deficits here also suggest impairments in self-other control, whereby self-other representations are poorly discriminated and self-representations, maladaptively inhibited.

1.2.2.1. The problem of the multimodal function of the TPJ. It is important to recognise that the TPJ has for a long time been described as a multimodal area of the brain, with the precise anatomy and functional role of this region remaining a controversial issue. For example, the TPJ has a well-documented role in the reorienting of attention (Corbetta & Shulman, 2002), the ability to switch one's focus of attention towards an unexpected, external stimulus. The TPJ has been considered a 'circuit breaker' in the context of attention (Corbetta, Patel, & Shulman, 2008). However, as already discussed, there is also a suggested role for the TPJ in various domains of social cognition (Decety & Lamm, 2007), and the more specific suggestion that it is involved in controlling or switching between representations of the self and of other people (Brass et al., 2005, 2009; Hogeveen et al., 2015; Santiesteban et al., 2012a, 2015a; Sowden & Catmur, 2015; Spengler et al., 2010b). Of particular contention is whether the region should be considered a unified region serving cognitive functions via a domain-general attentional control mechanism, or whether it has a domain-specific influence on each cognitive function it is reported to subserve (Donaldson, Rinehart, & Enticott, 2015). This is an important issue to consider with respect to the current thesis.

Despite claims that the TPJ cannot be selective for processes such as theory of mind and attentional orienting (Mitchell, 2008), recent evidence, as well as methodological flaws in the original investigation of this question, does indeed suggest something specialised about the TPJ's involvement in social cognition. For example, Scholz, Triantafyllou, Whitfield-Gabrieli, Brown, and Saxe (2009) claimed that previous findings of overlap in the precise

region of the TPJ recruited for attention and theory of mind (Mitchell, 2008) were due to low spatial resolution, and they in fact demonstrated distinct neural correlates for these two processes using higher resolution fMRI. Additionally, research is now revealing how the TPJ functions within different networks in the brain, and thus revealing other regions which might contribute to the neural input and output organisation of the TPJ. This is key in appreciating that no single brain region in isolation is responsible for a particular social cognitive function. Connectivity-based imaging has identified sub-regions within the TPJ, which participate in different cortical networks (Mars et al., 2012). An anterior cluster within the TPJ was found to have neural input and output projections to the ventral prefrontal cortex and anterior insula, while a posterior cluster of the TPJ was functionally connected to the posterior cingulate, temporal pole, and anterior medial prefrontal cortex. A recent activation likelihood estimation meta-analysis of 47 neuroimaging studies also confirms a role for the posterior right TPJ in the social domain, whilst anterior areas of the right TPJ were attributed to attentional as well as social processing (Krall et al., 2015). Thus, it is important to consider how the TPJ contributes to different functional networks which in combination supports its involvement in specific cognitive functions.

A handful of other issues must also be observed here, however. Firstly, the measures used in these studies to assess the TPJ's role in social vs. non-social processes often have very different cognitive demands and it is difficult to determine whether measures of social cognition are actually measuring something uniquely 'social' or just a domain general process. For example, it is difficult to rule out whether the differential neural responses found may simply reflect an aspect of the stimuli or task instructions, rather than differences in the underlying mechanism(s) of interest. Klapper, Ramsey, Wigboldus, and Cross (2014) manipulated the social nature of the imitation-inhibition task by modulating animacy of the observed hand stimulus within the same task. Utilising fMRI, they demonstrated increased

congruency effects in the right TPJ (indexed by greater differences in activation between congruent and incongruent trials) the more 'social' the task appeared to the participant (i.e. the more human-like the observed stimulus appeared to participants). This suggests the TPJ may be involved in more of a specialised capacity in imitation control, rather than simply a domain-general mechanism of conflict management. Nevertheless, there is still a lack of causal evidence for network distinctions within the TPJ. Sowden and Catmur (2015) sought to mitigate the above concerns by administering disruptive repetitive TMS to the right TPJ or a control mid-occipital (MO) site during a variation on the imitation-inhibition task. [2] They introduced a non-social, attentional control variant of the task, whereby performance can be measured both to control imitative response tendencies, but also to control the tendency to respond in a spatially compatible location to the observed finger movement. These two control processes (one social and one non-social in nature) have been shown to develop across different time courses (Catmur & Heyes, 2011), and Sowden and Catmur showed selective modulation of performance in the social, but not the non-social, control process of the task during disruptive repetitive TMS to the right TPJ.

In conclusion, evidence does support the idea that subregions within the TPJ may exist which participate in different neural networks and thus have the capacity to control different cognitive functions in a domain-specific fashion. Thus, investigating its involvement in a self-other control mechanism which is specific to social cognition seems well supported.

1.3. Atypical Social Cognitive Function

Uncovering a common low-level mechanism for social cognition centred around self-other representation seems particularly useful when considering atypical social cognitive function (Catmur, Cross, & Over, 2015). Mirror touch synaesthesia, a condition in which the

[2] The MO region was used as an active control site as there was no a priori reason to assume that stimulation to this region would differ from baseline.

observation of touch or pain to others elicits an overt somatic sensation in the synaesthete's own body, is associated with structural abnormalities in the TPJ, specifically with reduced grey matter volume (Grice-Jackson, Critchley, Banissy, & Ward, 2017; Holle, Banissy, & Ward, 2013). Behaviourally these individuals also demonstrate atypical self-other control, showing difficulties in imitation-inhibition (Santiesteban et al., 2015b), problems with judging self/other ownership over body parts (Aimola-Davies & White, 2013), and general atypical representation of the self (Maister, Banissy, & Tsakiris, 2013). Therefore, as individuals with mirror touch synaesthesia have abnormalities in the TPJ, in addition to those areas involved in vicarious somatosensory mirroring, and they have behavioural abnormalities in social domains requiring self-other control, mirror touch synaesthesia could be described as one example of a disorder of self-other control (Banissy & Ward, 2013; Santiesteban et al., 2015b; Ward & Banissy, 2015).

Furthermore, the ability to control neural representations of the self and of other people seems a central aspect of more common disorders of social cognition, such as autism and schizophrenia (Ferri et al., 2012; Spengler et al., 2010b). The aetiology and cognitive profile of ASD, the prototypical neurodevelopmental disorder of social cognition, will now be considered, and how the social impairments experienced by these individuals fit with the idea of impaired self-other control will be discussed.

1.3.1. Autism Spectrum Disorder and self-other control

1.3.1.1. Diagnosis and aetiology. Whilst originally considered to arise from environmental factors (Hanson & Gottesman, 1976), autism now stands out as one of the most highly heritable of all psychiatric and developmental disorders (Freitag, 2007; Plomin, Owen, & McGuffin, 1994). Behavioural genetics twin studies initially proposed conservative heritability estimates of 60% for autism as a single facet (Bailey et al., 1995; Folstein &

Rutter, 1977; Ritvo, Freeman, Mason-Brothers, Mo, & Ritvo, 1985; Steffenburg et al., 1989).[3] ASD is a life-long condition with an estimated prevalence of 1-2%, and is about three times more common in males than females (Baird et al., 2006; Baxter et al., 2015). However, it is suggested that females with ASD are often missed in the diagnosis process, perhaps because they are better able to compensate for their symptoms than males with ASD (Dworzynski, Ronald, Bolton, & Happé, 2012; Livingston & Happé, 2017).

Early research was quick to recognise that autism could be characterised by distinct impairments in behavioural functioning. This began with Leo Kanner's (Kanner, 1943; Kanner & Eisenberg, 1956) depiction of autism as a disorder with two main deficits; the social and the non-social and was later extended to form a diagnostically useful 'triad of impairments' (Figure 5) comprising social impairments, communication impairments and rigid and repetitive behaviours and interests (Rutter, 1978; Wing & Gould, 1978). Finally, in the most recent revision of the Diagnostic Statistical Manual-5, all subcategories (including Asperger's syndrome) were combined under the single heading of 'ASD' (American Psychiatric Association, 2013), and the triad of impairments was reduced to two categories; social communication and/or interaction difficulties, and restricted and/or repetitive behaviours. However, as it is still appreciated that ASD is a highly heterogenous disorder, a broad diagnosis is given along with details of the individual's specific areas of impairment.

Behavioural genetic studies also demonstrate that the core impairments of ASD are fractionable in their aetiology (Brunsdon & Happé, 2014). Heritability estimates increase to around 90% when social and non-social impairments are considered separately (Ronald & Hoekstra, 2011), and each have been shown to possess some separate genetic influences on

[3] Estimates of heritability are calculated by comparing concordance for behavioural traits between monozygotic and dizygotic twins. Concordance is calculated as the proportion of twins who possess the trait of interest and those who have a co-twin with the trait.

their development (Ronald, Happé, & Plomin, 2005; Ronald & Hoekstra, 2011).[4] Ronald and colleagues (Ronald et al., 2006) demonstrate strong genetic correlations between a single trait exhibited in the extremes of the population (the top 5% of the distribution) and the same trait exhibited sub-clinically in the rest of the distribution. This supports the notion that sub-threshold autistic traits not only exist, but possess the same aetiology throughout the population. It also supports the investigation of general population samples with respect to population level autistic traits to aid a greater understanding of autism throughout the spectrum. Thus, Chapters 2, 6 and 7 take a dimensional approach to the investigation of ASD traits/severity and social cognition.

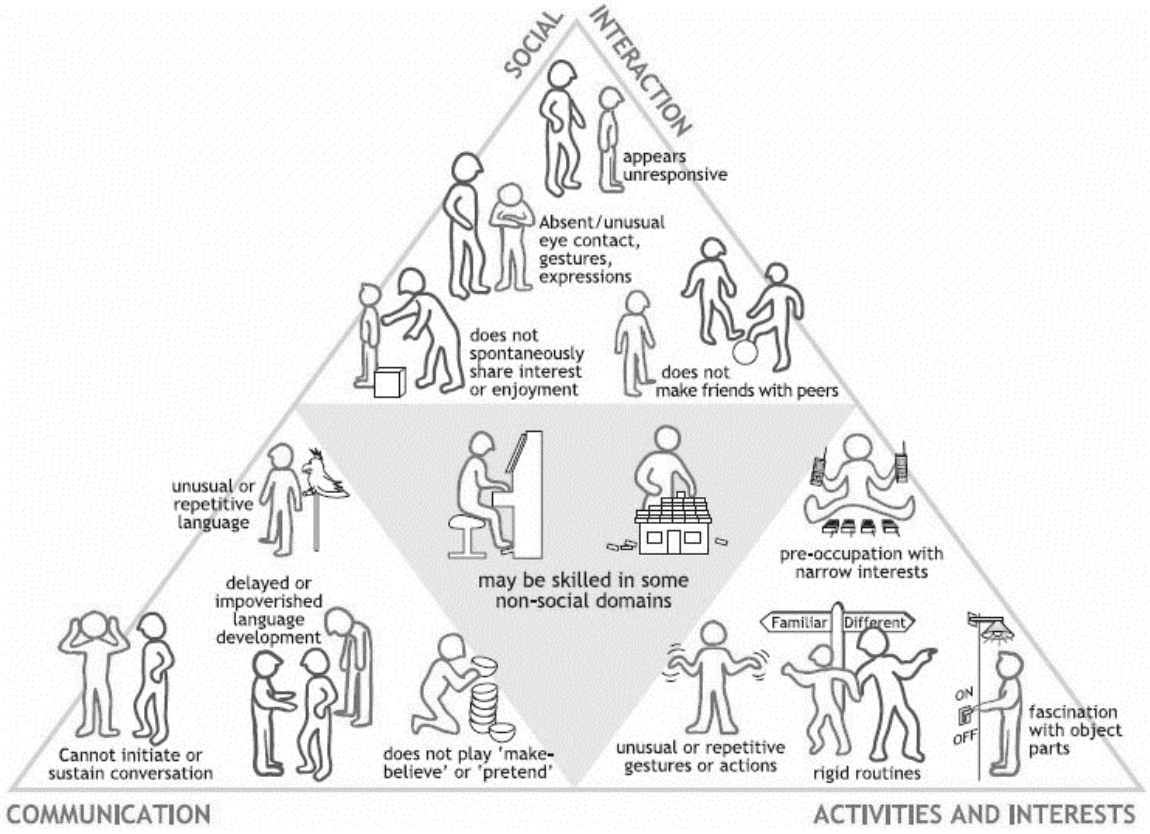

Figure 5. Visual depiction of the three main facets of the triadic impairments in autism suggested prior to the most recent revision of the Diagnostic Statistical Manual. Taken from http://labspace.open.ac.uk/mod/resource/view.php?id=482959.

[4] Multivariate genetic analyses examine genetic and environmental influences on the *covariance* between traits; the extent to which the genetic and environmental influences acting on one trait also act on one or more other traits.

In terms of the search for specific genes which contribute to the presence of autism, very few candidate genes with reliable and replicable effect sizes have been identified in the last three decades of molecular genetic research (Barnby & Monaco, 2003; Sykes & Lamb, 2007; Yang & Gill, 2007). It has in fact been proposed that multiple genes with small effect sizes may be partly responsible for many behavioural disorders such as ASD, which somewhat helps to explain the large heterogeneity in the autism phenotype.

1.3.1.2. Social cognitive impairments in ASD. As mentioned, autism is thought to be made up of two or three core impairments, including social communication and/or interaction, and rigid and/or repetitive behaviours and interests. Restrictive interests may involve preoccupation with particular classes of objects or with knowledge of a particular topic, as well as rigid routines (which induce anxiety if altered) and atypicalities in motor skills. The autism phenotype is also thought to include general poor executive functioning, which encompasses dysfunction in cognitive control such as planning, working memory, attention, problem solving, set-shifting and verbal reasoning (Courchesne et al., 1994; Happé, Booth, Charlton, & Hughes, 2006; Ozonoff, 1997) and is argued to reflect frontal lobe dysfunction (Hill, 2004). Weak central coherence, characterised by a focus on small details at the expense of the whole, is another non-social domain of impairment observed in individuals with ASD (Happé, 2005; Happé, Frith, & Briskman, 2001). Impairments in central coherence and executive functioning are suggested to account, in part, for the rigid and repetitive behaviours associated with ASD (Boyd, Mcbee, Holtzclaw, Baranek, & Bodfish, 2009; Kenworthy, Black, Harrison, Della Rosa, & Wallace, 2009; Le Monda, Holtzer, & Goldman, 2012), but also to relate to some aspects of social interaction (Kenworthy et al., 2009).

However, a prominent aspect of ASD is the associated broad range of social and communication impairments (Happé & Ronald, 2008; Rutter, 1978). Similar questions are now being asked regarding the inter-dependence of various cognitive impairments within the

social domain in ASD. If ASD is characterised by impairments across social cognitive functions found to be correlated in typically functioning individuals (Brass & Spengler, 2008; Brass et al., 2005, 2009; Spengler et al., 2009a), perhaps these may not only be developmentally linked, but may rely on a common underlying process (Happé et al., 2017; Sowden & Shah, 2014) which may be impaired in individuals with ASD. Therefore, although not an exhaustive outline of the cognitive impairments in ASD, an overview will be given of some of the social cognitive processes impaired in ASD which reveal the potential for a common parsimonious mechanistic explanation.

Individuals with autism consistently perform poorly during theory of mind tasks (Baron-Cohen, 2001; Baron-Cohen, Leslie, & Frith, 1985; Senju, Southgate, White, & Frith, 2009; Spengler et al., 2010b). These tasks require an individual to represent and distinguish mental states pertaining to both the self and others, including thoughts, desires and intentions, in order to predict behaviour (Premack & Woodruff, 1978). For example, children with autism often perform poorly on false belief tasks where they are required to understand that a character in a story may hold a different belief to oneself, or to another character in the story (Baron-Cohen, O'Riordan, Jones, Stone, & Plaisted, 1999b; Happé, 1994; White, Hill, Happé, & Frith, 2009). Many false beliefs tasks can be passed by high-functioning individuals with ASD (Bowler, 1992; Ozonoff, Pennington, & Rogers, 1991), however, once high enough demands are placed on theory of mind, high-functioning individuals with ASD do demonstrate impairments here (Castelli, Frith, Happé, & Frith, 2002; Mathersul, McDonald, & Rushby, 2013).

Furthermore, perspective-taking involves the ability to see the world from another person's perspective, and there are reports of atypical basic visual perspective-taking in individuals with autism (Dawson & Fernald, 1987; Hamilton, Brindley, & Frith, 2009; Leekam, Baron-Cohen, Perrett, Milders, & Brown, 1997; Reed, 2002; Warreyn, Roeyers,

Oelbrandt, & De Groote, 2005; Yirmiya, Sigman, & Zacks, 1994; Zwickel, White, Coniston, Senju, & Frith, 2010). However, there are some inconsistencies here, with not all studies finding deficits in ASD (see Pearson, Ropar, & Hamilton, 2013 for a review of this literature). There is some suggestion that this may be due to tasks in which ceiling effects are observed, particularly in high-functioning adult samples, as well as poor matching of intellectual function in a number of studies investigating perspective-taking in ASD (Pearson et al., 2013). Moreover, there is some discussion regarding different levels of perspective-taking which may exist, as well as whether or not perspective-taking can be considered an independent social cognitive construct to that of theory of mind (Hamilton et al., 2009).

Additionally, although not required diagnostic criteria for ASD, atypicalities have been reported in the ability to represent and distinguish emotional states and actions of the self and others, which may play a role in the ability to empathise with, and imitate, those with whom one interacts (Baron-Cohen & Wheelwright, 2004; Bird & Viding, 2014; Thaler et al., 2017; Williams, Whiten, & Singh, 2004; Williams, Whiten, Suddendorf, & Perrett, 2001). Importantly, however, these atypicalities are not always observed as *diminished* empathy or imitation. For example, there is suggestion that empathy may in fact be heightened in individuals with ASD (Markram, Rinaldi, & Markram, 2007; Smith, 2006, 2009), and it has also been reported that these individuals may not lack imitative behaviour, but may in fact over-imitate others' actions compared to controls (Bird, Leighton, Press, & Heyes, 2007; Spengler et al., 2010a). Thus, perhaps a mechanism of control or modulation of social behaviour may be the underlying deficit here. Furthermore, there is also the suggestion that deficits in empathy may not be characteristic of ASD, but may be associated with another condition called alexithymia, whereby individuals struggle to identify and describe their own and others' emotions (Bird & Cook, 2013; Cook, Brewer, Shah, & Bird, 2013; Sifneos, 1973), but we will return to the discussion of this later.

Despite the appealing proposition that these broad range of social impairments can be explained by a dysfunction in the MNS, termed the 'Broken Mirror Theory' of autism (Dapretto et al., 2006; Iacoboni & Dapretto, 2006; Ramachandran & Oberman, 2006; Rizzolatti & Fabbri-Desto, 2010; Williams et al., 2001, 2004), there is now support for an intact mirror system in individuals with ASD (Bird et al., 2007; Dinstein et al., 2010; Fan, Decety, Yang, Liu, & Cheng, 2010; Hamilton, Brindley, & Frith, 2007; Leighton, Bird, Charman, & Heyes, 2008; Press, Richardson, & Bird, 2010; Southgate & Hamilton, 2008; Spengler et al., 2010b). Furthermore, previous claims that mirror neurons are involved in everything from imitation (Iacoboni et al., 1999), to action understanding (Gallese & Sinigaglia, 2011; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996), theory of mind (Iacoboni et al., 2005), empathy (Avenanti, Bueti, Galati, & Aglioti, 2005) and language processing (Rizzolatti & Arbib, 1998) are now highly disputed (Cook, Bird, Catmur, Press, & Heyes, 2014), and thus the Broken Mirror Theory seems a less plausible common explanation for the social impairments observed in ASD.

Nevertheless, it is important to consider that ASD may still be characterised by atypical neural representations of the self and of other people even if not in the form of atypical MNS function. Various investigations have revealed atypical neural self-representation in prefrontal regions of the brain (Kennedy & Courchesne, 2008; Lombardo et al., 2010b; Uddin et al., 2008) as well as reduced neural representation of others (Hobson & Meyer, 2005; Kennedy & Courchesne, 2008; Meyer & Hobson, 2004). A handful of studies have postulated deficient self-other differentiation or control to explain impairments in inhibiting imitation, theory of mind and perspective-taking (Lombardo, Chakrabarti, Bullmore, & Baron-Cohen, 2011; Lombardo et al., 2010b; Spengler et al., 2010a, 2010b). The latter studies in particular deserve more attention. Lombardo and colleagues (Lombardo et al., 2011) identified impairments in the recruitment of the right TPJ when making

judgments requiring self-other differentiation in individuals with ASD relative to controls. Thus, although it is clear that both mPFC and right TPJ are crucially involved in neural representations of the self and the other (Brass et al., 2005, 2009), the right TPJ appears particularly important in facilitating the *control* or modulation of these competing representations. Similarly, Spengler and colleagues (Spengler et al., 2010b) found that, in a sample of high functioning individuals with ASD, increased imitation was associated with reduced theory of mind and decreased activity in areas typically required for self-other control. Despite varied terminology, including self-other "differentiation," "distinction," "switching" or "agency," all postulated processes appear to share a common feature of the "control" of shared representations.

Indeed, more recently it has been suggested that the processes deficient in ASD are those involved in the top-down modulation of social behaviour (Cook, Barbalat, & Blakemore, 2012; Cook & Bird, 2012; Southgate & Hamilton, 2008; Spengler et al., 2010b; Wang & Hamilton, 2012). Of particular note, Cook and Bird (2012) found that the modulatory effects of priming pro-social attitudes on self-other control observed in neurotypical adults were absent in individuals with ASD. Moreover, individuals with ASD have been shown to respond much more slowly to the effects of being imitated on enhancements in empathy for others' pain, when compared to typical controls (De Coster, Wiersema, Deschrijver, & Brass, 2017).

Consequently, the mechanism of self-other control considered here, as a form of top-down modulation of behaviour, has the capacity to explain the broad range of social deficits commonly observed in ASD. Naturally this may also extend to our understanding of the social impairments observed in various other clinical disorders such as schizophrenia. Schizophrenia is a disorder also associated with behavioural impairments in the control of self and other representations (Ferri et al., 2012; Graham-Schmidt, Martin-Iverson, & Waters,

in press; van der Weiden, Prikken, & van Haren, 2015), as well as abnormalities in the structure and function of the TPJ (Allen, Larøi, McGuire, & Aleman, 2008; Benedetti et al., 2009; Brüne et al., 2008, 2011; Cook et al., 2012). Taken together, neuroimaging evidence to date is suggestive of the TPJ's important role in a common neurocognitive mechanism which may be deficient across disorders. Let us now consider a further condition which we know to be highly co-morbid with ASD and which is suggested to account for the emotional symptoms of ASD (Bird & Cook, 2013); alexithymia.

1.3.2. Alexithymia and self-other control. Alexithymia is a sub-clinical construct characterised by an inability to identify and describe one's own and others' emotions (Sifneos, 1973) and it is now known to co-occur with a range of psychiatric disorders associated with social cognitive impairments (Brewer, Cook, & Bird, 2016a; Murphy, Brewer, Catmur, & Bird, 2017). These include ASD (Bird & Cook, 2013; Cook et al., 2013), eating disorders (Brewer, Cook, Cardi, Treasure, & Bird, 2015), schizophrenia (Van't Wout, Aleman, Bermond, & Kahn, 2007), addiction (Verdejo-Garcia, Clark, & Dunn, 2012), obsessive compulsive disorder (Grabe et al., 2006), anxiety (Hendryx, Haviland, & Shaw, 1991), and major depressive disorder (Honkalampi, Hintikka, Tanskanen, Lehtonen, & Viinamäki, 2000). For example, the prevalence of alexithymia in the general population is 5-10%, but approximately 50% in the ASD population (Berthoz & Hill, 2005; Hill, Berthoz, & Frith, 2004; Kokkonen et al., 2001) and 60% in the eating disorder population (Cochrane, Brewerton, Wilson, & Hodges, 1993). Alexithymia has now been suggested to be characterised as a general impairment in interoception; the perception of the internal state of one's own body (Brewer, Cook, & Bird, 2016a, 2016b; Murphy, Catmur, & Bird, in press), and recent computational models within the predictive coding framework suggest that the ability to distinguish between self and other is also dependent upon interoception (Seth, 2013; Quattrocki & Friston, 2014).

As it is suggested that good interoceptive awareness may be crucial for awareness of one's own body and its distinction from others, it may be conceptually predicted that lower interoceptive accuracy (and accordingly, high levels of alexithymia) would result in a reduced ability to control competing representations of the self and others (Quattrocki & Friston, 2014). However, one empirical study in fact found that superior interoceptive accuracy was related to greater difficulties inhibiting the tendency to imitate (Ainley, Brass, & Tsakiris, 2014). Nevertheless, it seems intuitive that alexithymia may indeed be characterised by atypicalities with self-other control, but it is an open question as to the specificity of the relationship between alexithymia, interoception and self-other control related processing. The relationship between alexithymia and self-other control will be investigated in Chapter 3 of the current thesis.

1.4. Conclusions and Open Questions

Thus far, behavioural and neuroscientific evidence for self-other control as a candidate neurocognitive mechanism for social cognition has been explored. Its involvement in key domains of social functioning, such as imitative control, theory of mind, perspective-taking and empathy has been outlined, with the suggestion that the mechanism may provide a resolution to paradoxes of social interaction, whereby in some situations (such as imitation-inhibition) the self must be enhanced and the other inhibited, whilst in other domains the self must be inhibited and the other enhanced (such as theory of mind, perspective-taking and empathy). Furthermore, a clear neural basis of this mechanism has been outlined in the context of imitation-inhibition and perspective-taking, two domains for which self-other control appears important. Finally, although we do not yet know the precise developmental trajectories for the neurocognitive deficits observed in disorders such as ASD, schizophrenia, and mirror touch synaesthesia, the survey of the existing literature suggests that consideration

of a self-other control mechanism may explain the broad sociocognitive deficits seen in these conditions.

However, there are a number of outstanding questions concerning the behavioural and neural bases of self-other control, as well as its precise involvement in atypical as well as typical social cognitive functioning, which the current thesis seeks to address. First, with the exception of a select few studies, investigations of self-other representation to date have suffered from methodological issues. For example, the investigation of one's ability to represent the self and the other in disorders of social cognition have often been done via independent tasks, and comparisons are then made at the neural level between these tasks (Langdon et al., 1997). Accordingly, future assessment of self-other control in disorders of social cognition can benefit from the now widely used task of imitative control (Figure 1) as a robust behavioural index, amalgamating the inspection of both self- and other-representations in one task. All chapters in this thesis take the approach of investigating the simultaneous measurement of self and other representations and this thesis seeks a greater understanding of whether self-other control might also explain deficits across *domains* of social cognition as well as across *disorders* of social cognition.

Chapters 2 and 3 will begin by examining the prototypical index of self-other control (the 'control of imitation', 'automatic imitation' or 'imitation-inhibition' task) in relation to disorders of social cognition. Chapter 2 (Sowden et al., 2016a) considers the integrity of the Broken Mirror Theory as an explanation of social functioning deficits in autism, and whether ASD may be better characterised as a disorder of control or modulation of social behaviour. Chapter 3 (Sowden et al., 2016b) examines the specificity of any link between self-other control and alexithymia. Both Chapters 2 and 3 also have the capacity to shed light on the question of domain-generality vs domain-specificity of this mechanism in explaining imitative control in these disorders.

In order to validate self-other control as a mechanism branching across social domains, there is a need for more behavioural measures of self-other control which elicit similar demands of self-other conflict across different domains of social cognition. There is also an open question as to the degree to which this mechanism may extend to more high-level social cognition. Therefore, Chapter 5 (Sowden et al., 2015), Experiment 1 discusses the development of a novel task to index self-other control in the context of lie detection (where representations here relate to the opinions of self and other which must be controlled for successful lie detection performance). Experiment 2 of this chapter then tests whether the lie detection task developed in Experiment 1 is affected by excitatory TPJ stimulation in the same way as other tasks assessing self-other control. If so, one can be more confident of the TPJ's role in self-other control across domains of social cognition. This is preceded by Chapter 4 which provides an overview of the history, physiology and methodology of tDCS. Following these chapters, Chapter 6 further examines the integrity of self-other control in relation to ASD utilising the lie detection self-other control task. This chapter examines self-other control in relation to autistic traits distributed throughout the population (Trembath & Vivanti, 2014).

Finally, Chapter 7 describes the development and piloting of a novel group task to elicit self-other interference effects in the context of social conformity; a higher-level, group social process, in which the 'other' refers not simply to one individual but to a group of others. It is currently an open and important question to consider how a mechanism of self-other control may apply to group social interaction. Moreover, there is the suggestion that there may exist two varieties of social influence; acceptance and compliance, with the reasons for social conformity here concerning whether or not one's response in a task is private or public to the rest of the group. Thus, this chapter also aims to dissociate these processes

within the same task to afford the investigation of two types of self-other conflict within the same social domain.

Chapter 7 is written with a focus on task design and methodology, with supplementary consideration of how self-other interference effects in the domain of conformity relate to population level autistic traits. This acts as a primer for future research using this task to further understand the neural and behavioural bases of self-other control.

1.5. References

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nature Reviews Neuroscience, 4,* 165-178.

Aimola-Davies, A. M., & White, R. C. (2013). A sensational illusion: Vision-touch synaesthesia and the rubber hand paradigm. *Cortex, 49,* 806-818.

Ainley, V., Brass, M., & Tsakiris, M. (2014). Heartfelt imitation: High interoceptive awareness is linked to greater automatic imitation. *Neuropsychologia, 60,* 21–28.

Allen, P., Larøi, F., McGuire, P. K., & Aleman, A. (2008). The hallucinating brain: A review of structural and functional neuroimaging studies of hallucinations. *Neuroscience & Biobehavioral Reviews, 32,* 175-191.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorder* (5th ed.). Arlington, VA: American Psychiatric Association.

Anderson, S. W., Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1999). Impairment of social and moral behavior related to early damage in human prefrontal cortex. *Nature Neuroscience, 2,* 1032-1037.

Avenanti, A., Bueti, D., Galati, G. & Aglioti, S. M. (2005) Transcranial magnetic stimulation highlights the sensorimotor side of empathy for pain. *Nature Neuroscience, 8,* 955–960.

Bailey, A., Le Couteur, A., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, E., & Rutter, M. (1995). Autism as a strongly genetic disorder: Evidence from a British twin study. *Psychological Medicine, 25,* 63-77.

Baird, G., Simonoff, E., Pickles, A., Chandler, S., Loucas, T., Meldrum, D., & Charman, T. (2006). Prevalence of disorders of the autism spectrum in a population cohort of

children in South Thames: The Special Needs and Autism Project (SNAP). *The Lancet, 368,* 210–215.

Banissy, M. J., & Ward, J. (2013). Mechanisms of self-other representations and vicarious experiences of touch in mirror-touch synesthesia. *Frontiers in Human Neuroscience, 7*, 112.

Bardi, L., Six, P., & Brass, M. (2017). Repetitive TMS of the temporo-parietal junction disrupts participant's expectations in a spontaneous Theory of Mind task. *Social Cognitive and Affective Neuroscience*, nsx109. doi.org/10.1093/scan/nsx109.

Barnby, G., & Monaco, A. P. (2003). Strategies for autism candidate gene analysis. In G. Bock & J. Goode (Eds.), *Autism: Neural basis and treatment possibilities. Novartis Foundation Symposium* 251 (pp. 48 – 63). Chichester, UK: John Wiley.

Baron-Cohen, S. (2001). Theory of mind in normal development and autism. *Prisme, 34*, 74-183.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"?. *Cognition, 21,* 37-46.

Baron-Cohen, S., O'Riordan, M., Jones, R., Stone, V., & Plaisted, K. (1999b). A new test of social sensitivity: Detection of faux pas in normal children and children with Asperger syndrome. *Journal of Autism and Developmental Disorders, 29,* 407–418

Baron- Cohen, S., Ring, H. A., Wheelwright, S., Bullmore, E. T., Brammer, M. J., Simmons, A., & Williams, S. C. (1999a). Social intelligence in the normal and autistic brain: An fMRI study. *European Journal of Neuroscience, 11*, 1891-1898.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*, 163-175.

Baxter, S. J., Brugha, T. S., Erskine, H. E., Scheurer, R. W., Vos, T., & Scott, J. G. (2015). The epidemiology and global burden of autism spectrum disorders. *Psychological Medicine*, *45*, 601–613.

Benedetti, F., Bernasconi, A., Bosia, M., Cavallaro, R., Dallaspezia, S., Falini, A., ... & Smeraldi, E. (2009). Functional and structural brain correlates of theory of mind and empathy deficits in schizophrenia. *Schizophrenia Research*, *114*, 154-160.

Berthoz, S., & Hill, E. L. (2005). The validity of using self-reports to assess emotion regulation abilities in adults with autism spectrum disorder. *European Psychiatry*, *20*, 291–298.

Bird, G., & Cook, R. (2013). Mixed emotions: The contribution of alexithymia to the emotional symptoms of autism. *Translational Psychiatry*, *3*, e285.

Bird, G., Leighton, J., Press, C., & Heyes, C. (2007). Intact automatic imitation of human and robot actions in autism spectrum disorders. *Proceedings of the Royal Society B: Biological Sciences*, *274*, 3027-3031.

Bird, G., & Viding, E. (2014). The self to other model of empathy: Providing a new framework for understanding empathy impairments in psychopathy, autism, and alexithymia. *Neuroscience & Biobehavioral Reviews*, *47*, 520-532.

Blakemore, S. J., Winston, J., & Frith, U. (2004). Social cognitive neuroscience: Where are we heading?. *Trends in Cognitive Sciences*, *8*, 216-222.

Botvinick, M., Jha, A. P., Bylsma, L. M., Fabian, S. A., Solomon, P. E., & Prkachin, K. M. (2005). Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain. *NeuroImage*, *25*, 312-319.

Boyd, B. A., Mcbee, M., Holtzclaw, T., Baranek, G. T., & Bodfish, J. W. (2009). Relationships among repetitive behaviors, sensory features, and executive functions in high functioning autism. *Research in Autism Spectrum Disorders*, *3*, 959–966.

Bowler, D. M. (1992). "Theory of Mind" in Asperger's Syndrome. *Journal of Child Psychology and Psychiatry*, *33*, 877–893.

Brass, M., Bekkering, H., & Prinz, W. (2001). Movement observation affects movement execution in a simple response task. *Acta Psychologica*, *106*, 3-22.

Brass, M., Bekkering, H., Wohlschläger, A., & Prinz, W. (2000). Compatibility between observed and executed finger movements: Comparing symbolic, spatial, and imitative cues. *Brain and Cognition*, *44*, 124-143.

Brass, M., Derrfuss, J., Matthes-von Cramon, G., & von Cramon, D. Y. (2003). Imitative response tendencies in patients with frontal brain lesions. *Neuropsychology*, *17*, 265-271.

Brass, M., Derrfuss, J., & von Cramon, D. Y. (2005). The inhibition of imitative and overlearned responses: A functional double dissociation. *Neuropsychologia*, *43*, 89-98.

Brass, M., & Heyes, C. (2005). Imitation: Is cognitive neuroscience solving the correspondence problem? *Trends in Cognitive Sciences*, *9*, 489-495.

Brass, M., Ruby, P., & Spengler, S. (2009). Inhibition of imitative behaviour and social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*, 2359-2367.

Brass, M., & Spengler, S. (2008). The inhibition of imitative behavior and the attribution of mental states. In T. Striano & V. Reid (Eds.), *Social cognition: Development, neuroscience and autism* (pp. 52-66). Oxford, UK: Blackwell.

Brewer, R., Cook, R., & Bird, G. (2016a). Shared interoceptive representations: The case of alexithymia. In S. S. Obhi & E. S. Cross (Eds.), *Shared representations: Sensorimotor foundations of social life*. Cambridge, UK: Cambridge University Press.

Brewer, R., Cook, R., & Bird, G. (2016b). Alexithymia: A general deficit of interoception. *Royal Society Open Science, 3*, 150664.

Brewer, R., Cook, R., Cardi, V., Treasure, J., & Bird, G. (2015). Emotion recognition deficits in eating disorders are explained by co-occurring alexithymia. *Royal Society Open Science, 2*, 140382

Brüne, M., Lissek, S., Fuchs, N., Witthaus, H., Peters, S., Nicolas, V., ... & Tegenthoff, M. (2008). An fMRI study of theory of mind in schizophrenic patients with "passivity" symptoms. *Neuropsychologia*, *46*, 1992-2001.

Brüne, M., Özgürdal, S., Ansorge, N., von Reventlow, H. G., Peters, S., Nicolas, V., ... & Lissek, S. (2011). An fMRI study of "theory of mind" in at-risk states of psychosis: comparison with manifest schizophrenia and healthy controls. *NeuroImage, 55*, 329-337.

Brunet, E., Sarfati, Y., Hardy-Baylé, M. C., & Decety, J. (2000). A PET investigation of the attribution of intentions with a nonverbal task. *NeuroImage, 11*, 157-166.

Brunet, E., Sarfati, Y., Hardy-Baylé, M. C., & Decety, J. (2003). Abnormalities of brain function during a nonverbal theory of mind task in schizophrenia. *Neuropsychologia, 41*, 1574-1582.

Brunsdon, V. E., & Happé, F. (2014). Exploring the 'fractionation' of autism at the cognitive level. *Autism, 18*, 17-30.

Calder, A. J., Keane, J., Manes, F., Antoun, N., & Young, A. W. (2000). Impaired recognition and experience of disgust following brain injury. *Nature Neuroscience, 3*, 1077-1078.

Castelli, F., Frith, C., Happé, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain, 125*, 1839-1849.

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

Catmur, C., Cross, E. S., & Over, H. (2016). Understanding self and others: From origins to disorders. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*, 20150066.

Catmur, C., & Heyes, C. (2011). Time course analyses confirm independence of imitative and spatial compatibility. *Journal of Experimental Psychology: Human Perception and Performance, 37*, 409-421.

Cazzato, V., Mian, E., Serino, A., Mele, S., & Urgesi, C. (2015). Distinct contributions of extrastriate body area and temporoparietal junction in perceiving one's own and others' body. *Cognitive, Affective, & Behavioral Neuroscience, 15*, 211-228.

Chaminade, T., & Decety, J. (2002). Leader or follower? Involvement of the inferior parietal lobule in agency. *Neuroreport, 13,* 1975-1978.

Chaminade, T., Meltzoff, A. N., & Decety, J. (2002). Does the end justify the means? A PET exploration of the mechanisms involved in human imitation. *NeuroImage, 15,* 318-328.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology, 76,* 893-910.

Chartrand, T. L., Cheng, C. M., & Jefferis, V. E. (2002). You're just a chameleon: The automatic nature and social significance of mimicry. In M. Jarymowicz & R. Ohme (Eds.), *Nature of automaticity*. (pp. 19-24). Warsaw, PL: Polish Academy of Sciences & Social Psychology.

Cochrane, C. E., Brewerton, T. D., Wilson, D. B., & Hodges, E. L. (1993). Alexithymia in the eating disorders. *The International Journal of Eating Disorders, 14,* 219–222.

Cook, J. L. (2014). Task-relevance dependent gradients in medial prefrontal and temporoparietal cortices suggest solutions to paradoxes concerning self/other control. *Neuroscience & Biobehavioral Reviews, 42,* 298-302.

Cook, J., Barbalat, G., & Blakemore, S. J. (2012). Top-down modulation of the perception of other people in schizophrenia and autism. *Frontiers in Human Neuroscience, 6, 175.*

Cook, J., & Bird, G. (2011). Social attitudes differentially modulate imitation in adolescents and adults. *Experimental Brain Research, 211,* 601-612.

Cook, J. L., & Bird, G. (2012). Atypical social modulation of imitation in autism spectrum conditions. *Journal of Autism and Developmental Disorders, 42,* 1045-1051.

Cook, R., Bird, G., Catmur, C., Press, C., & Heyes, C. (2014). Mirror neurons: From origin to function. *Behavioral and Brain Sciences, 37,* 177-192.

Cook, R., Brewer, R., Shah, P., & Bird, G. (2013). Alexithymia, not autism, predicts poor recognition of emotional facial expressions. *Psychological Science, 24,* 723–732.

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: From environment to theory of mind. *Neuron, 58,* 306-324.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience, 3,* 201-215.

Costa, A., Torriero, S., Oliveri, M., & Caltagirone, C. (2008). Prefrontal and temporo-parietal involvement in taking others' perspective: TMS evidence. *Behavioural Neurology, 19,* 71-74.

Courchesne, E., Townsend, J., Akshoomoff, N. A., Saitoh, O., Yeung-Courchesne, R., Lincoln, A. J., ... & Lau, L. (1994). Impairment in shifting attention in autistic and cerebellar patients. *Behavioral Neuroscience, 108,* 848-865.

Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y., & Iacoboni, M. (2006). Understanding emotions in others: Mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience, 9,* 28-30.

David, N., Bewernick, B. H., Cohen, M. X., Newen, A., Lux, S., Fink, G. R., ... & Vogeley, K. (2006). Neural representations of self versus other: Visual-spatial perspective taking and agency in a virtual ball-tossing game. *Journal of Cognitive Neuroscience, 18,* 898-910.

Dawson, G., & Fernald, M. (1987). Perspective-taking ability and its relationship to the social behavior of autistic children. *Journal of Autism and Developmental Disorders, 17,* 487-498.

Decety, J., & Chaminade, T. (2003). Neural correlates of feeling sympathy. *Neuropsychologia, 41*, 127-138.

Decety, J., Chaminade, T., Grezes, J., & Meltzoff, A. N. (2002). A PET exploration of the neural mechanisms involved in reciprocal imitation. *NeuroImage, 15,* 265-272.

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

Decety, J., & Sommerville, J. A. (2003). Shared representations between self and other: A social cognitive neuroscience view. *Trends in Cognitive* Sciences, 7, 527-533.

De Coster, L., Wiersema, J. R., Deschrijver, E., & Brass, M. (2017). The effect of being imitated on empathy for pain in adults with high-functioning autism: Disturbed self–other distinction leads to altered empathic responding. *Autism*. doi.org/10.1177/1362361317701268.

de Guzman, M., Bird, G., Banissy, M. J., & Catmur, C. (2016). Self–other control processes in social cognition: From imitation to empathy. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*, 20150079.

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

Dinstein, I., Thomas, C., Humphreys, K., Minshew, N., Behrmann, M., & Heeger, D. J. (2010). Normal movement selectivity in autism. *Neuron, 66,* 461-469.

di Pellegrino, G. D., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research*, *91*, 176-180.

Donaldson, P. H., Rinehart, N. J., & Enticott, P. G. (2015). Noninvasive stimulation of the temporoparietal junction: A systematic review. *Neuroscience & Biobehavioral Reviews*, *55*, 547-572.

Dworzynski, K., Ronald, A., Bolton, P., & Happé, F. (2012). How different are girls and boys above and below the diagnostic threshold for autism spectrum disorders?. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*, 788-797.

Fan, Y. T., Decety, J., Yang, C. Y., Liu, J. L., & Cheng, Y. (2010). Unbroken mirror neurons in autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, *51*, 981-988.

Farrer, C., & Frith, C. D. (2002). Experiencing oneself vs another person as being the cause of an action: The neural correlates of the experience of agency. *NeuroImage*, *15*, 596-603.

Farrer, C., Franck, N., Frith, C. D., Decety, J., Georgieff, N., d'Amato, T., & Jeannerod, M. (2004). Neural correlates of action attribution in schizophrenia. *Psychiatry Research: Neuroimaging*, *131*, 31-44.

Farrer, C., Franck, N., Georgieff, N., Frith, C. D., Decety, J., & Jeannerod, M. (2003). Modulating the experience of agency: A positron emission tomography study. *NeuroImage*, *18*, 324-333.

Feinberg, T. E., Venneri, A., Simone, A. M., Fan, Y., & Northoff, G. (2010). The neuroanatomy of asomatognosia and somatoparaphrenia. *Journal of Neurology, Neurosurgery and Psychiatry*, *81*, 276-281.

Ferri, F., Frassinetti, F., Mastrangelo, F., Salone, A., Ferro, F. M., & Gallese, V. (2012). Bodily self and schizophrenia: The loss of implicit self-body knowledge. *Consciousness and Cognition*, *21*, 1365-1374.

Ferstl, E. C., & von Cramon, D. Y. (2002). What does the frontomedian cortex contribute to language processing: Coherence or theory of mind?. *NeuroImage*, *17*, 1599-1612.

Fletcher, P. C., Happe, 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*, 109-128.

Folstein, S. E., & Rutter, M. (1977). Genetic influences and infantile autism. *Nature*, 265, 726–728.

Freitag, C. M. (2007). The genetics of autistic disorders and its clinical relevance: A review of the literature. *Molecular Psychiatry*, *12*, 2-22.

Frith, U., & Frith, C. (2010). The social brain: Allowing humans to boldly go where no other species has been. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*, 165-176.

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

Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: An fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia*, *38*, 11-21.

Gallese, V., & Sinigaglia, C. (2011). What is so special about embodied simulation?. *Trends in Cognitive Sciences, 15*, 512-519.

Grabe, H. J., Ruhrmann, S., Ettelt, S., Müller, A., Buhtz, F., Hochrein, A., ... & Pukrop, R. (2006). Alexithymia in obsessive-compulsive disorder–results from a family study. *Psychotherapy and Psychosomatics, 75*, 312-318.

Graham-Schmidt, K. T., Martin-Iverson, M. T., & Waters, F. A. (in press). Self-and other-agency in people with passivity (first rank) symptoms in schizophrenia. *Schizophrenia Research*. doi.org/10.1016/j.schres.2017.04.024.

Grèzes, J., Frith, C. D., & Passingham, R. E. (2004). Inferring false beliefs from the actions of oneself and others: An fMRI study. *NeuroImage, 21*, 744-750.

Grice-Jackson, T., Critchley, H. D., Banissy, M. J., & Ward, J. (2017). Common and distinct neural mechanisms associated with the conscious experience of vicarious pain. *Cortex, 94*, 152-163.

Hamilton, A. F. D. C., Brindley, R. M., & Frith, U. (2007). Imitation and action understanding in autistic spectrum disorders: How valid is the hypothesis of a deficit in the mirror neuron system?. *Neuropsychologia, 45*, 1859-1868.

Hamilton, A. F. D. C., Brindley, R., & Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition, 113*, 37-44.

Hanson, D. R., & Gottesman, I. I. (1976). The genetics, if any, of infantile autism and childhood schizophrenia. *Journal of Autism and Childhood Schizophrenia, 6*, 209-234.

Happé, F.G. (1994). An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*, 129–154.

Happé, F. (2005). The weak central coherence account of autism. In F. R. Volkmar, R. Paul, A. Klin & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders, diagnosis, neurobiology, and behavior* (3rd ed.) (Vol. 1) (pp. 640-649). New Jersey, NJ: John Wiley & Sons.

Happé, F., Booth, R., Charlton, R., & Hughes, C. (2006). Executive function deficits in autism spectrum disorders and attention-deficit/hyperactivity disorder: Examining profiles across domains and ages. *Brain and Cognition*, *61*, 25–39.

Happé, F., Cook, J. L., & Bird, G. (2017). The structure of social cognition: In (ter) dependence of sociocognitive processes. *Annual Review of Psychology*, *68*, 243-267.

Happé, F., Frith, U., & Briskman, J. (2001). Exploring the cognitive phenotype of autism: Weak "central coherence" in parents and siblings of children with autism: I. Experimental tests. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*, 299-307.

Happé, F., & Ronald, A. (2008). The 'fractionable autism triad': A review of evidence from behavioural, genetic, cognitive and neural research. *Neuropsychology Review*, *18*, 287-304.

Harris, P. L., de Rosnay, M., & Pons, F. (2005). Language and children's understanding of mental states. *Current Directions in Psychological Science*, *14*, 69-73.

Hendryx, M. S., Haviland, M. G., & Shaw, D. G. (1991). Dimensions of alexithymia and their relationships to anxiety and depression. *Journal of Personality Assessment, 56*, 227-237.

Hennenlotter, A., Schroeder, U., Erhard, P., Castrop, F., Haslinger, B., Stoecker, D., ... & Ceballos-Baumann, A. O. (2005). A common neural basis for receptive and expressive communication of pleasant facial affect. *NeuroImage, 26*, 581-591.

Heyes, C., Bird, G., Johnson, H., & Haggard, P. (2005). Experience modulates automatic imitation. *Cognitive Brain Research, 22*, 233-240.

Hill, E. L. (2004). Evaluating the theory of executive dysfunction in autism. *Developmental Review, 24*, 189-233.

Hill, E., Berthoz, S., & Frith, U. (2004). Brief report: Cognitive processing of own emotions in individuals with autistic spectrum disorder and in their relatives. *Journal of Autism and Developmental Disorders, 34*, 229–235.

Hobson, R. P., & Meyer, J. A. (2005). Foundations for self and other: A study in autism. *Developmental Science, 8*, 481-491.

Hogeveen, J., Chartrand, T. L., & Obhi, S. S. (2014). Social mimicry enhances mu-suppression during action observation. *Cerebral Cortex, 25*, 2076-2082.

Hogeveen, J., & Obhi, S. S. (2012). Social interaction enhances motor resonance for observed human actions. *The Journal of Neuroscience, 32*, 5984-5989.

Hogeveen, J., Obhi, S. S., Banissy, M. J., Santiesteban, I., Press, C., Catmur, C., & Bird, G. (2015). Task-dependent and distinct roles of the temporoparietal junction and inferior frontal cortex in the control of imitation. *Social Cognitive and Affective Neuroscience, 10*, 1003-1009.

Holle, H., Banissy, M. J., & Ward, J. (2013). Functional and structural brain differences associated with mirror-touch synaesthesia. *NeuroImage, 83*, 1041–1050.

Honkalampi, K., Hintikka, J., Tanskanen, A., Lehtonen, J., & Viinamäki, H. (2000). Depression is strongly associated with alexithymia in the general population. *Journal of Psychosomatic Research, 48*, 99-104.

Hutchison, W. D., Davis, K. D., Lozano, A. M., Tasker, R. R., & Dostrovsky, J. O. (1999). Pain-related neurons in the human cingulate cortex. *Nature Neuroscience, 2*, 403-405.

Hynes, C. A., Baird, A. A., & Grafton, S. T. (2006). Differential role of the orbital frontal lobe in emotional versus cognitive perspective-taking. *Neuropsychologia, 44*, 374-383.

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

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C. & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology*, *3*, 529–535.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. S*cience, 286*, 2526-2528.

Jackson, P. L., Brunet, E., Meltzoff, A. N., & Decety, J. (2006). Empathy examined through the neural mechanisms involved in imagining how I feel versus how you feel pain. *Neuropsychologia, 44*, 752-761.

Jackson, P. L., Meltzoff, A. N., & Decety, J. (2005). How do we perceive the pain of others? A window into the neural processes involved in empathy. *NeuroImage, 24*, 771-779.

Kable, J. W., & Chatterjee, A. (2006). Specificity of action representations in the lateral occipitotemporal cortex. *Journal of Cognitive Neuroscience, 18*, 1498-1517.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child, 2*, 217-250.

Kanner, L., & Eisenberg, L. (1956). Early infantile autism 1943–1955. *American Journal of Orthopsychiatry*, *26*, 55–65.

Kennedy, D. P., & Courchesne, E. (2008). Functional abnormalities of the default network during self-and other-reflection in autism. *Social Cognitive and Affective Neuroscience*, *3*, 177-190.

Kenworthy, L., Black, D. O., Harrison, B., Della Rosa, A., & Wallace, G. L. (2009). Are executive control functions related to autism symptoms in high functioning children? *Child Neuropsychology*, *15*, 425–440.

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.

Keysers, C., & Gazzola, V. (2009). Expanding the mirror: Vicarious activity for actions, emotions, and sensations. *Current Opinion in Neurobiology*, *19*, 666-671.

Klapper, A., Ramsey, R., Wigboldus, D., & Cross, E. S. (2014). The control of automatic imitation based on bottom-up and top-down cues to animacy: Insights from brain and behavior. *Journal of Cognitive Neuroscience*, *26*, 2503-2513.

Kokkonen, P., Karvonen, J. T., Veijola, J., Läksy, K., Jokelainen, J., Järvelin, M. -R., & Joukamaa, M. (2001). Prevalence and sociodemographic correlates of alexithymia in a population sample of young adults. *Comprehensive Psychiatry*, *42*, 471–476.

Krall, S. C., Rottschy, C., Oberwelland, E., Bzdok, D., Fox, P. T., Eickhoff, S. B., ... & Konrad, K. (2015). The role of the right temporoparietal junction in attention and social interaction as revealed by ALE meta-analysis. *Brain Structure and Function, 220*, 587-604.

Lakin, J. L., Jefferis, V. E., Cheng, C. M., & Chartrand, T. L. (2003). The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior, 27*, 145-162.

Lamm, C., Batson, C. D., & Decety, J. (2007). The neural substrate of human empathy: Effects of perspective-taking and cognitive appraisal. *Journal of Cognitive Neuroscience, 19*, 42-58.

Lamm, C., Decety, J., & Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *NeuroImage, 54*, 2492-2502.

Langdon, R., Michie, P. T., Ward, P. B., McConaghy, N., Catts, S. V., & Coltheart, M. (1997). Defective self and/or other mentalising in schizophrenia: A cognitive neuropsychological approach. *Cognitive Neuropsychiatry, 2*, 167-193.

Lawrence, E. J., Shaw, P., Giampietro, V. P., Surguladze, S., Brammer, M. J., & David, A. S. (2006). The role of 'shared representations' in social perception and empathy: An fMRI study. *NeuroImage, 29*, 1173-1184.

Leekam, S. R., Baron-Cohen, S., Perrett, D., Milders, M., & Brown, S. D. (1997). Eye-direction detection: A dissociation between geometric and joint attention skills in autism. *British Journal of Developmental Psychology, 15*, 77-95.

Leighton, J., Bird, G., Charman, T., & Heyes, C. (2008). Weak imitative performance is not due to a functional 'mirroring' deficit in adults with Autism Spectrum Disorders. *Neuropsychologia, 46*, 1041-1049.

Leighton, J., Bird, G., Orsini, C., & Heyes, C. (2010). Social attitudes modulate automatic imitation. *Journal of Experimental Social Psychology, 46*, 905-910.

LeMonda, B. C., Holtzer, R., & Goldman, S. (2012). Relationship between executive functions and motor stereotypies in children with autistic disorder. *Research in Autism Spectrum Disorders*, *6*, 1099–1106.

Leube, D. T., Knoblich, G., Erb, M., Grodd, W., Bartels, M., & Kircher, T. T. (2003). The neural correlates of perceiving one's own movements. *NeuroImage*, *20*, 2084-2090.

Livingston, L. A., & Happé, F. (2017). Conceptualising compensation in neurodevelopmental disorders: Reflections from autism spectrum disorder. *Neuroscience & Biobehavioral Reviews*, *80*, 729-742.

Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., & Baron-Cohen, S. (2011). Specialization of right temporo-parietal junction for mentalizing and its relation to social impairments in autism. *NeuroImage, 56*, 1832-1838.

Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., Sadek, S. A., Pasco, G., Wheelwright, S. J., ... & Baron-Cohen, S. (2010b). Atypical neural self-representation in autism. *Brain*, *133*, 611-624.

Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., Wheelwright, S. J., Sadek, S. A., Suckling, J., & Baron-Cohen, S. (2010a). Shared neural circuits for mentalizing about the self and others. *Journal of Cognitive Neuroscience*, *22*, 1623-1635.

Mai, X., Zhang, W., Hu, X., Zhen, Z., Xu, Z., Zhang, J., & Liu, C. (2016). Using tDCS to explore the role of the right temporo-parietal junction in theory of mind and cognitive empathy. *Frontiers in Psychology*, *7*, 380.

Maister, L., Banissy, M. J., & Tsakiris, M. (2013). Mirror-touch synaesthesia changes representations of self-identity. *Neuropsychologia*, *51*, 802-808.

Markram, H., Rinaldi, T., & Markram, K. (2007). The intense world syndrome: An alternative hypothesis for autism. *Frontiers in Neuroscience*, *1*, 77-96.

Mars, R. B., Sallet, J., Schüffelgen, U., Jbabdi, S., Toni, I., & Rushworth, M. F. (2012). Connectivity-based subdivisions of the human right "temporoparietal junction area": evidence for different areas participating in different cortical networks. *Cerebral Cortex*, *22*, 1894-1903.

Mathersul, D., McDonald, S., & Rushby, J. A. (2013). Understanding advanced theory of mind and empathy in high-functioning adults with autism spectrum disorder. *Journal of Clinical and Experimental Neuropsychology*, *35*, 655-668.

Meyer, J. A., & Hobson, R. P. (2004). Orientation in relation to self and other: The case of autism. *Interaction Studies*, *5*, 221-244.

Mitchell, J. P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cerebral Cortex*, *18*, 262-271.

Molenberghs, P., Cunnington, R., & Mattingley, J. B. (2012). Brain regions with mirror properties: A meta-analysis of 125 human fMRI studies. *Neuroscience & Biobehavioral Reviews*, *36*, 341-349.

Moriguchi, Y., Decety, J., Ohnishi, T., Maeda, M., Mori, T., Nemoto, K., ... & Komaki, G. (2007). Empathy and judging other's pain: An fMRI study of alexithymia. *Cerebral Cortex*, *17*, 2223-2234.

Moriguchi, Y., Ohnishi, T., Lane, R. D., Maeda, M., Mori, T., Nemoto, K., ... & Komaki, G. (2006). Impaired self-awareness and theory of mind: An fMRI study of mentalizing in alexithymia. *NeuroImage*, *32*, 1472-1482.

Moskowitz, G. B. (2005). *Social cognition: Understanding self and others*. New York, NY: Guilford Press.

Murphy, J., Brewer, R., Catmur, C., & Bird, G. (2017). Interoception and psychopathology: A developmental neuroscience perspective. *Developmental Cognitive Neuroscience*, *23*, 45-56.

Murphy, J., Catmur, C., & Bird, G. (in press). Alexithymia is associated with a multi-domain, multi-dimensional failure of interoception: Evidence from novel tests. *Journal of Experimental Psychology*: *General*.

Nobusako, S., Nishi, Y., Shuto, T., Asano, D., Osumi, M., & Morioka, S. (2017). Transcranial direct current stimulation of the temporoparietal junction and inferior frontal cortex improves imitation-inhibition and perspective-taking with no effect on the Autism-Spectrum Quotient score. *Frontiers in Behavioral Neuroscience*, *11*, 84.

Obhi, S. S., & Cross, E. S. (Eds.). (2016). *Shared Representations: Sensorimotor Foundations of Social Life*. Cambridge, UK: Cambridge University Press.

Obhi, S. S., & Hogeveen, J. (2013). The controlled imitation task: a new paradigm for studying self-other control. *PeerJ, 1*, e161.

Ohnishi, T., Moriguchi, Y., Matsuda, H., Mori, T., Hirakata, M., Imabayashi, E., ... & Uno, A. (2004). The neural network for the mirror system and mentalizing in normally developed children: An fMRI study. *Neuroreport*, *15*, 1483-1487.

Ozonoff, S. (1997). Components of executive function in autism and other disorders. In J. Russell (Ed.), *Autism as an executive disorder* (pp. 179–211). New York, NY: Oxford University Press.

Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991). Executive function deficits in high-functioning autistic individuals: Relationship to theory of mind. *Journal of Child Psychology and Psychiatry*, *32*, 1081–1105.

Pearson, A., Ropar, D., & Hamilton, A. F. D. C. (2013). A review of visual perspective taking in autism spectrum disorder. *Frontiers in Human Neuroscience*, *7*, 652.

Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Thinking of mental and other representations: The roles of left and right temporo-parietal junction. *Social Neuroscience*, *1*, 245-258.

Plomin, R., Owen, M. J., & McGuffin, P. (1994). The genetic basis of complex human behaviors. *Science*, *264*, 1733-1739.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and Brain Sciences*, *1*, 515-526.

Press, C., Richardson, D., & Bird, G. (2010). Intact imitation of emotional facial actions in autism spectrum conditions. *Neuropsychologia*, *48*, 3291-3297.

Quattrocki, E., & Friston, K. (2014). Autism, oxytocin and interoception. *Neuroscience & Biobehavioral Reviews*, *47*, 410–430.

Ramachandran, V. S., Blakeslee, S. (1998). *Phantoms in the brain: Probing the mysteries of the human mind*. New York, NY: William Morrow.

Ramachandran, V. S., & Oberman, L. M. (2006). Broken mirrors: A theory of autism. *Scientific American, 295,* 62-69.

Ramnani, N., & Miall, R. C. (2004). A system in the human brain for predicting the actions of others. *Nature Neuroscience, 7,* 85-90.

Reed, T. (2002). Visual perspective taking as a measure of working memory in participants with autism. *Journal of Developmental and Physical Disabilities, 14,* 63-76.

Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2004). The neural correlates of theory of mind within interpersonal interactions. *NeuroImage, 22,* 1694-1703.

Ritvo, E. R., Freeman, B. J., Mason-Brothers, A., Mo, A., & Ritvo, A. M. (1985). Concordance for the syndrome of autism in 40 pairs of afflicted twins. *American Journal of Psychiatry, 142,* 74–77.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences, 21,* 188-194.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27,* 169-192.

Rizzolatti, G., & Fabbri-Destro, M. (2010). Mirror neurons: From discovery to autism. *Experimental Brain Research, 200,* 223-237.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3,* 131-141.

Ronald, A., Happé, F., Bolton, P., Butcher, L. M., Price, T. S., Wheelwright, S., et al. (2006). Genetic heterogeneity between the three components of the autism spectrum: A twin study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *45*, 691–699.

Ronald, A., Happé, F., & Plomin, R. (2005). The genetic relationship between individual differences in social and nonsocial behaviours characteristic of autism. *Developmental Science*, *8*, 444–458.

Ronald, A., & Hoekstra, R. A. (2011). Autism spectrum disorders and autistic traits: A decade of new twin studies. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *156*, 255-274.

Ruby, P., & Decety, J. (2001). Effect of subjective perspective taking during simulation of action: A PET investigation of agency. *Nature Neuroscience*, *4*, 546-550.

Ruby, P., & Decety, J. (2003). What you believe versus what you think they believe: A neuroimaging study of conceptual perspective- taking. *European Journal of Neuroscience*, *17*, 2475-2480.

Ruby, P., & Decety, J. (2004). How would you feel versus how do you think she would feel? A neuroimaging study of perspective-taking with social emotions. *Journal of Cognitive Neuroscience*, *16*, 988-999.

Rutter, M. (1978). Diagnosis and definition of childhood autism. *Journal of Autism and Childhood Schizophrenia*, *8*, 139–161.

Santiesteban, I., Banissy, M. J., Catmur, C., & Bird, G. (2012a). Enhancing social ability by stimulating right temporoparietal junction. *Current Biology*, *22*, 2274-2277.

Santiesteban, I., Banissy, M. J., Catmur, C., & Bird, G. (2015a). Functional lateralization of temporoparietal junction–imitation inhibition, visual perspective- taking and theory of mind. *European Journal of Neuroscience*, *42*, 2527-2533.

Santiesteban, I., Bird, G., Tew, O., Cioffi, M. C., & Banissy, M. J. (2015b). Mirror-touch synaesthesia: Difficulties inhibiting the other. *Cortex*, *71*, 116-121.

Santiesteban, I., White, S., Cook, J., Gilbert, S. J., Heyes, C., & Bird, G. (2012b). Training social cognition: From imitation to theory of mind. *Cognition*, *122*, 228-235.

Saxe, R. (2009). Theory of mind (neural basis). In W. Banks (Ed.), *Encyclopedia of consciousness* (pp. 401–410). Oxford, UK: Academic Press.

Saxe, R. (2010). The right temporo-parietal junction: A specific brain region for thinking about thoughts. In A. Leslie & T. German (Eds.), *Handbook of theory of mind*. Hove, UK: Psychology Press.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind". *NeuroImage*, *19*, 1835-1842.

Saxe, R., Moran, J. M., Scholz, J., & Gabrieli, J. (2006). Overlapping and non-overlapping brain regions for theory of mind and self reflection in individual subjects. *Social Cognitive and Affective Neuroscience*, *1*, 229-234.

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.

Saxe, R., & Wexler, A. (2005). Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia*, *43*, 1391-1399.

Saxe, R., Xiao, D. K., Kovacs, G., Perrett, D. I., & Kanwisher, N. (2004). A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologica, 42*, 1435-1446.

Schnell, K., Heekeren, K., Schnitker, R., Daumann, J., Weber, J., Heßelmann, V., ... & Gouzoulis-Mayfrank, E. (2007). An fMRI approach to particularize the frontoparietal network for visuomotor action monitoring: Detection of incongruence between test subjects' actions and resulting perceptions. *NeuroImage*, *34*, 332-341.

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.

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: An absence of spontaneous theory of mind in Asperger syndrome. *Science*, *325*, 883-885.

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences, 17,* 565–573.

Sifneos, P. E. (1973). The prevalence of "alexithymic" characteristics in psychosomatic patients. *Psychotherapy and Psychosomatics, 22*, 255–262.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science, 303*, 1157-1162.

Smith, A. (2006). Cognitive empathy and emotional empathy in human behaviour and evolution. *Psychological Record, 56*, 3-21.

Smith, A. (2009). The empathy imbalance hypothesis of autism: A theoretical approach to cognitive and emotional empathy in autistic development. *Psychological Record, 59*, 273-294.

Southgate, V., & de C Hamilton, A. F. (2008). Unbroken mirrors: Challenging a theory of autism. *Trends in Cognitive Sciences, 12*, 225-229.

Sowden, S., Brewer, R., Catmur, C., & Bird, G. (2016b). The specificity of the link between alexithymia, interoception, and imitation. *Journal of Experimental Psychology: Human Perception and Performance, 42*, 1687-1692.

Sowden, S., & Catmur, C. (2015). The role of the right temporoparietal junction in the control of imitation. *Cerebral Cortex, 25*, 1107-1113.

Sowden, S., Koehne, S., Catmur, C., Dziobek, I., & Bird, G. (2016a). Intact automatic imitation and typical spatial compatibility in autism spectrum disorder: Challenging the broken mirror theory. *Autism Research, 9*, 292-300.

Sowden, S., & Shah, P. (2014). Self-other control: A candidate mechanism for social cognitive function. *Frontiers in Human Neuroscience, 8*, 789.

Sowden, S., Wright, G. R., Banissy, M. J., Catmur, C., & Bird, G. (2015). Transcranial current stimulation of the temporoparietal junction improves lie detection. *Current Biology, 25*, 2447-2451.

Spence, S. A., Brooks, D. J., Hirsch, S. R., Liddle, P. F., Meehan, J., & Grasby, P. M. (1997). A PET study of voluntary movement in schizophrenic patients experiencing passivity phenomena (delusions of alien control). *Brain, 120*, 1997-2011.

Spengler, S., Bird, G., & Brass, M. (2010a). Hyperimitation of actions is related to reduced understanding of others' minds in autism spectrum conditions. *Biological Psychiatry, 68*, 1148-1155.

Spengler, S., von Cramon, D. Y., & Brass, M. (2009a). Control of shared representations relies on key processes involved in mental state attribution. *Human Brain Mapping, 30*, 3704-3718.

Spengler, S., von Cramon, D. Y., & Brass, M. (2009b). Was it me or was it you? How the sense of agency originates from ideomotor learning revealed by fMRI. *NeuroImage, 46*, 290-298.

Spengler, S., von Cramon, D. Y., & Brass, M. (2010b). Resisting motor mimicry: Control of imitation involves processes central to social cognition in patients with frontal and temporo-parietal lesions. *Social Neuroscience, 5*, 401-416.

Sperduti, M., Delaveau, P., Fossati, P., & Nadel, J. (2011). Different brain structures related to self-and external-agency attribution: A brief review and meta-analysis. *Brain Structure and Function, 216*, 151-157.

Steffenburg, S., Gillberg, C., Hellgren, L., Andersson, L., Gillberg, I. C., Jakobsson, G., & Bohman, M. (1989). A twin study of autism in Denmark, Finland, Iceland, Norway and Sweden. *Journal of Child Psychology and Psychiatry, 30*, 405-416.

Sturmer, B., Aschersleben, G., & Prinz, W. (2000). Correspondence effects with manual gestures and postures: A study of imitation. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 1746–1759.

Sykes, N. H., & Lamb, J. A. (2007). Autism: The quest for the genes. *Expert Reviews in Molecular Medicine, 9*, 1-15.

Tallal, P. (2004). Improving language and literacy is a matter of time. *Nature Reviews Neuroscience, 5,* 721-728.

Thaler, H., Gebauer, L., Skewes, J. C., Christensen, P., Prkachin, K. M., & Jegindø Elmholdt, E. M. (2017). Typical pain experience but underestimation of others' pain: Emotion perception in self and others in autism spectrum disorder. *Autism*. doi.org/10.1177/1362361317701269.

Trembath, D., & Vivanti, G. (2014). Problematic but predictive: individual differences in children with autism spectrum disorders. *International Journal of Speech-Language Pathology, 16,* 57-60.

Tsakiris, M., Costantini, M., & Haggard, P. (2008). The role of the right temporo-parietal junction in maintaining a coherent sense of one's body. *Neuropsychologia, 46,* 3014-3018.

Uddin, L. Q., Davies, M. S., Scott, A. A., Zaidel, E., Bookheimer, S. Y., Iacoboni, M., & Dapretto, M. (2008). Neural basis of self and other representation in autism: An FMRI study of self-face recognition. *PloS one, 3,* e3526.

van der Weiden, A., Prikken, M., & van Haren, N. E. (2015). Self–other integration and distinction in schizophrenia: A theoretical analysis and a review of the evidence. *Neuroscience & Biobehavioral Reviews, 57,* 220-237.

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

Van't Wout, M., Aleman, A., Bermond, B., & Kahn, R. S. (2007). No words for feelings: Alexithymia in schizophrenia patients and first-degree relatives. *Comprehensive Psychiatry, 48,* 27-33.

Verdejo-Garcia, A., Clark, L., & Dunn, B. D. (2012). The role of interoception in addiction: A critical review. *Neuroscience & Biobehavioral Reviews, 36,* 1857-1869.

Voellm, B. A., Taylor, A. N., Richardson, P., Corcoran, R., Stirling, J., McKie, S., … & Elliot, R. (2006). Neuronal correlates of theory of mind and empathy: A functional magnetic resonance imaging study in a nonverbal task. *NeuroImage, 29,* 90-98.

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happe, F., Falkai, P., ... & Zilles, K. (2001). Mind reading: Neural mechanisms of theory of mind and self-perspective. *NeuroImage, 14,* 170-181.

Walter, H., Adenzato, M., Ciaramidaro, A., Enrici, I., Pia, L., & Bara, B. G. (2004). Understanding intentions in social interaction: The role of the anterior paracingulate cortex. *Journal of Cognitive Neuroscience, 16,* 1854-1863.

Wang, Y., & Hamilton, A. F. D. C. (2012). Social top-down response modulation (STORM): A model of the control of mimicry in social interaction. *Frontiers in Human Neuroscience, 6,* 153.

Ward, J., & Banissy, M. J. (2015). Explaining mirror-touch synesthesia. *Cognitive Neuroscience, 6,* 118-133.

Warreyn, P., Roeyers, H., Oelbrandt, T., & De Groote, I. (2005). What are you looking at? Joint attention and visual perspective taking in young children with autism spectrum disorder. *Journal of Developmental and Physical Disabilities, 17,* 55-73.

White, S., Hill, E., Happe, F., & Frith, U. (2009). Revisiting the strange stories: Revealing mentalizing impairments in autism. *Child Development, 80,* 1097–1117.

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron, 40*, 655-664.

Williams, J. H., Waiter, G. D., Gilchrist, A., Perrett, D. I., Murray, A. D., & Whiten, A. (2006). Neural mechanisms of imitation and 'mirror neuron' functioning in autistic spectrum disorder. *Neuropsychologia*, *44*, 610-621.

Williams, J. H., Whiten, A., & Singh, T. (2004). A systematic review of action imitation in autistic spectrum disorder. *Journal of Autism and Developmental Disorders*, *34*, 285-299.

Williams, J. H., Whiten, A., Suddendorf, T., & Perrett, D. I. (2001). Imitation, mirror neurons and autism. *Neuroscience & Biobehavioral Reviews, 25*, 287-295.

Wing, L., & Gould, J. (1978). Systematic recording of behaviors and skills of retarded and psychotic children. *Journal of Autism and Childhood Schizophrenia*, *8*, 79–97.

Yang, M. S., & Gill, M. (2007). A review of gene linkage, association and expression studies in autism and an assessment of convergent evidence. *International Journal of Developmental Neuroscience, 25*, 69-85.

Yirmiya, N., Sigman, M., & Zacks, D. (1994). Perceptual perspective-taking and seriation abilities in high-functioning children with autism. *Development and Psychopathology, 6*, 263-272.

Young, L., Camprodon, J. A., Hauser, M., Pascual-Leone, A., & Saxe, R. (2010). Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proceedings of the National Academy of Sciences, 107*, 6753-6758.

Young, L., Cushman, R., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences, 104*, 8235-8240.

Zwickel, J., White, S. J., Coniston, D., Senju, A., & Frith, U. (2010). Exploring the building blocks of social cognition: Spontaneous agency perception and visual perspective taking in autism. *Social Cognitive and Affective Neuroscience*, *6*, 564-571.

Chapter 2. Intact automatic imitation and typical spatial compatibility in autism spectrum disorder: Challenging the broken mirror theory

This chapter is presented as a published article and is an exact copy of the following journal publication:

Sowden, S.*, Koehne, S.*, Catmur, C., Dziobek, I., & Bird, G. (2016). Intact automatic imitation and typical spatial compatibility in autism spectrum disorder: Challenging the broken mirror theory. *Autism Research*, *9*, 292-300.

*Joint first author

Corresponding Author:

Sophie Sowden, MRC Social, Genetic and Developmental Psychiatry, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, SE5 8AF, United Kingdom.

sophie.sowden@kcl.ac.uk

RESEARCH ARTICLE

Intact Automatic Imitation and Typical Spatial Compatibility in Autism Spectrum Disorder: Challenging the Broken Mirror Theory

Sophie Sowden, Svenja Koehne, Caroline Catmur, Isabel Dziobek, and Geoffrey Bird

A lack of imitative behavior is frequently described as a core feature of Autism Spectrum Disorder (ASD), and is consistent with claims of mirror neuron system dysfunction in these individuals. Previous research has questioned this characterization of ASD however, arguing that when tests of automatic imitation are used—which do not require higher-level cognitive processing—imitative behavior is intact or even enhanced in individuals with ASD. In Experiment 1, 60 adult individuals with ASD and a matched Control group completed an automatic imitation task in which they were required to perform an index or a middle finger lift while observing a hand making either the same, or the alternate, finger movement. Both groups demonstrated a significant imitation effect whereby actions were executed faster when preceded by observation of the same action, than when preceded by the alternate action. The magnitude of this "imitation effect" was statistically indistinguishable in the ASD and Control groups. Experiment 2 utilized an improved automatic imitation paradigm to demonstrate that, when automatic imitation effects are isolated from those due to spatial compatibility, increasing autism symptom severity is associated with an increased tendency to imitate. Notably, there was no association between autism symptom severity and spatial compatibility, demonstrating the specificity of the link between ASD symptoms and increased imitation. These results provide evidence against claims of a lack of imitative behavior in ASD, and challenge the "Broken Mirror Theory of Autism." ***Autism Res** 2016, 9: 292–300.* © 2015 International Society for Autism Research, Wiley Periodicals, Inc.

Keywords: autism; imitation; mirror neurons; broken mirror theory; individual differences

Introduction

Autism Spectrum Disorder (ASD) is defined by persistent difficulties in social communication and interaction, with accompanying restricted and repetitive patterns of thought and behavior [American Psychiatric Association, 2013]. The social and non-social impairments associated with ASD have been argued to be due to different aetiological factors, making a single explanation for the symptoms associated with ASD unlikely [Happé, Ronald, & Plomin, 2006]. Within the social domain a great deal of attention has been focussed on the ability of individuals with ASD to imitate, due to the theorized importance of imitation for the development of socio-cognitive abilities including empathy, theory of mind, and language [Rogers & Pennington, 1991]. The focus on imitation in ASD has been further increased due to the claim that an atypical or "broken" mirror neuron system (MNS), the neural system subserving imitation [Catmur, Walsh, & Heyes, 2009; Cook, Bird, Catmur, Press, & Heyes, 2014; Heiser, Iacoboni, Maeda, Marcus, & Mazziotta, 2003; Iacoboni et al., 1999], may be responsible for the symptoms of ASD [Ramachandran & Oberman, 2006].

The Broken Mirror Theory has prompted a great deal of research examining the structural and functional integrity of the MNS in autism (see Hamilton, 2013 for a review of these studies). Empirical evidence examining the Broken Mirror Theory has produced highly mixed results, with as many studies reporting typical MNS structure and function in ASD as those finding impairments [Hamilton, 2013]. Of concern is the fact that several studies claiming to support the Broken Mirror Theory have examined "mu" suppression using Electroencephalograpy (EEG), which has recently been shown to index sensory rather than motor features of observed actions, and therefore to be unsuitable as an index of MNS functioning [Coll, Bird, Catmur, & Press, 2015]. Other techniques used to measure MNS functioning in humans have also been criticized [Hamilton,

Sophie Sowden and Svenja Koehne are joint first authors
From the MRC Social, Genetic and Developmental Psychiatry, Institute of Psychiatry, Psychology and Neuroscience, King's College London (S.S., G.B.); Department of Psychology, Humboldt University, Berlin (S.K., I.D.); Department of Psychology, University of Surrey, Surrey GU2 7XH, United Kingdom (C.C.); Institute of Cognitive Neuroscience, University College London, London WC1N 3AR, United Kingdom (G.B.)
Received March 26, 2015; accepted for publication May 20, 2015
Address for correspondence and reprints: Sophie Sowden, Department for Sophie Sowden and Geoff Bird is MRC Social, Genetic and Developmental Psychiatry, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London SE5 8AF, United Kingdom. Email: sophie.sowden@kcl.ac.uk
Published online 26 June 2015 in Wiley Online Library (wileyonlinelibrary.com)
DOI: 10.1002/aur.1511

2013; Southgate & Hamilton, 2008], making imitation perhaps the best current index of MNS function.

Williams, Whiten, and Singh [2004] provided the first systematic review of imitative behavior in ASD, selecting 21 studies judged to be the most methodologically robust of an original 121 investigations. 14 studies found atypical imitative behavior in ASD, while the remaining 7 were judged to have produced uninterpretable results, prompting the authors to the overall conclusion that there is evidence of a specific imitation impairment in autism. It is noteworthy however, that all of the reviewed studies measured intentional or *voluntary* imitation—where the participant is explicitly asked to copy a model action [e.g. the "do-as-I-do" task, Hayes & Hayes, 1952]. These typically involve a number of higher-level cognitive and motivational processes known to be impaired in individuals with ASD, such as executive functioning, pragmatic language understanding, and rapid attentional switching [Courchesne et al., 1994; Happé & Frith, 1995; Russell, 1997; Williams, Goldstein, Carpenter, & Minshew, 2005]. Thus, it may be impairments in processes other than imitation which lead to poor performance in tasks of voluntary imitation [Bird, Leighton, Press, & Heyes, 2007]. Indeed, Leighton, Bird, Charman, and Heyes [2008] demonstrated that individuals with ASD were as impaired on a nonimitative version of a voluntary "do-as-I-do" task as they were on the imitative version. This result challenges the notion that ASD is associated with a *specific* imitation impairment—suggesting instead that poor performance on tests of voluntary imitation is due to more *general* cognitive impairments. It has therefore been suggested that in order to obtain a "pure" measure of imitative performance in ASD one must use automatic imitation tasks, due to their reduced reliance on abilities other than imitation [Bird et al., 2007].

Tests of automatic imitation, based on classic Stimulus-Response Compatibility (SRC) paradigms [Kornblum, Hasbroucq, & Osman, 1990; Prinz, 1997], require action responses to be made to action stimuli, where the specific action performed by the stimulus is task-irrelevant. Stürmer, Aschersleben, and Prinz [2000] first utilized this paradigm to demonstrate that individuals are faster to execute a hand movement whilst observing execution of the same (compatible) rather than a different (incompatible) hand movement. Such tasks are thought to measure automatic rather than intentional imitation because the movement compatibility is formally task-irrelevant yet significantly influences task responses.

Most recent versions of the automatic imitation paradigm [Brass et al., 2000, 2005, 2009; Catmur & Heyes, 2011; Catmur et al., 2009; Gowen, Bradshaw, Galpin, Lawrence, & Poliakoff, 2010; Leighton & Heyes, 2010] utilize a choice reaction-time (RT) task in which a non-action cue (a number or colored shape) specifies both the onset and type of response required for each trial (i.e. the cue acts as discriminative stimulus and imperative cue). For example, Brass and colleagues [Brass et al., 2000, 2005, 2009] instructed participants to perform an index or middle finger lift when presented with the number "1" or "2," respectively. At the same time, a task-irrelevant stimulus hand lifted either the same (imitatively compatible trial), or the alternative (imitatively incompatible trial) finger to that lifted by the participant. RT measures are then used to assess automatic imitation, where imitatively compatible trials result in speeded, and imitatively incompatible trials in slowed, responses when compared with baseline trials on which no action is observed.

A handful of studies have investigated automatic imitation in this manner in individuals with ASD and found it to be either typical [Hamilton, Brindley, & Frith, 2007; Press, Richardson, & Bird, 2010; Schunke et al., 2015], or even enhanced [Bird et al., 2007; Spengler, Bird, & Brass, 2010]. Findings of increased automatic imitation is consistent with clinical symptoms of echolalia and echopraxia (the involuntary copying of the speech and actions of others) in ASD [Grossi, Marcone, Cinquegrana, & Gallucci, 2012]. However, research to date has been restricted to small samples, potentially under-powered to detect group differences, and has included a number of studies utilizing facial emotion expressions or complex joint action tasks [McIntosh, Reichmann-Decker, Winkielman, & Wilbarger, 2006; Press et al., 2010; Sebanz, Knoblich, Stumpf, & Prinz, 2005]. Consequently, it could be argued that task performance is more a product of face, emotion, or higher level processing abilities than action imitation.

Experiment 1

Experiment 1 sought to provide a robust replication of previous findings of intact automatic imitation in a large sample of individuals with ASD using affect-neutral finger actions. An adapted version of the automatic imitation task [see Cook & Bird, 2012], originally presented by Brass et al. [2000], was used in this study. The task provides two dissociable measures; automatic imitation (AI), the degree to which observation of an action (e.g. lifting of the index finger) prompts performance of the same action, and effector compatibility (EC), the nonimitative tendency for any response made with an effector to be executed faster when cued by the same effector, than when cued by a different effector. On EC trials, the on-screen hand remained stationary, rather than performing a lifting movement as on AI trials, and instead the compatible or incompatible finger (effector) was highlighted with a semitransparent green mask. EC trials provide a nonimitative control, as they draw attention to the effector and influence response times, without movement of the finger, and therefore

Table 1. Participant Characteristics for Experiment 1 and 2

	Experiment 1		Experiment 2	
	ASD	Control	ASD	Control
N	60	45	18	18
Gender	39 Male, 21 Female	26 Male, 19 Female	16 Male, 2 Female	15 Male, 3 Female
Mean Age (years)	33.3 (*9.2*)	32.5 (*9.7*)	37.7 (*12.6*)	34.9 (*15.2*)
Mean Full-scale IQ	109.2 (*15.2*)	108.2 (*12.5*)	113.6 (*13.8*)	114.7 (*10.8*)
Mean AQ	36.1 (*8.6*)	13.7 (*4.6*)	35.1 (*9.0*)	17.1 (6.3)
ADOS Classification	25 Autism, 18 Autism Spectrum	n/a	11 Autism, 7 Autism Spectrum	n/a
Mean ADOS-G Score	9.4 (*0.5*)	n/a	10.61 (*2.7*)	n/a

NB. ADOS classifications and mean ADOS-G scores apply only to the 43 ASD individuals who met the ADOS classification cutoff ($\geq$7).

without the observed action necessary for imitation [Leighton & Heyes, 2010].

Method

Participants. Sixty high-functioning individuals with a clinical diagnosis of ASD (39 male; $M_{AGE} = 32.3$ years; SD = 9.2) were recruited through the autism outpatient clinic of the Charité University Medicine Berlin or referred by specialized clinicians and centers. An age-, gender-, and IQ-matched sample of 45 healthy control individuals (26 male; $M_{AGE} = 32.5$ years; SD = 9.7) was recruited via online advertisements and a mailing list for potential participants held at the Freie Universität Berlin. All participants had normal or corrected-to-normal vision and individuals with ASD were diagnosed according to DSM-IV criteria [American Psychiatric Association, 1994], which was confirmed using the Autism Diagnostic Observation Schedule [ADOS; Lord et al., 2000]. Twelve individuals in the ASD group, with a clinical diagnosis of ASD, failed to later meet the ADOS Autism Spectrum cutoff (score of 7 or above) and a further five did not have an available ADOS score. The pattern of results was not altered by the inclusion or exclusion of these individuals, and thus the results reported below include all 60 cases with a clinical diagnosis of ASD.

Autism Spectrum Quotient [AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001] scores were significantly higher [$t(103) = 15.8$, $P < 0.001$] in the ASD group ($M = 36.1$, SD = 8.6) than the Control group ($M = 13.7$, SD = 4.6), and the groups did not differ significantly in age [$t(103) = 0.41$, $P = 0.680$], proportion of females [$\chi^2(1) = 0.57$, $P = 0.45$], or IQ [$t(103) = 0.33$, $P = 0.741$]. Full details of ASD and Control group characteristics are presented in Table 1.

Stimuli and Procedure. The task was that used by Cook and Bird [2012] and the stimuli consisted of short video clips (subtending visual angles of 6° vertically and 9° horizontally and of 3000 ms in duration) of a stimulus hand presented on a blue background and rotated around the sagittal and transverse planes with respect to the participant's hand (Fig. 1A).

Participants were seated approximately 80 cm from the computer screen with their right arm supported on the table in front of them, and required throughout the task to use the index and middle fingers of their right hand to depress the "V" and "B" keys on the computer keyboard. On each trial participants were required to lift either their index or their middle finger on the appearance of a "1" or "2," respectively. On 50% of trials participants simultaneously viewed a five-frame video clip of a human hand lifting the index or middle finger. This lifting action was either compatible or incompatible with the participant's required finger lift (AI manipulation). On the remaining 50% of trials participants viewed three-frame trials on which the fingers of the observed hand remained static but either the compatible or incompatible finger was covered by a green mask (EC manipulation).

Participants first completed practice trials and were required to make five consecutive correct responses for each trial type (AI and EC). The main experiment consisted of 120 trials presented in a pseudo-random order, with an approximate duration of 15 min.

Results

Participants who were observed to be significant outliers (>1.5 times the interquartile range of AI or EC effects for each group) were excluded prior to data analysis (1 Control, 2 ASD). As in Cook and Bird [2011, 2012], trials in which an inaccurate finger-lift response was made (5% of trials for Control group; 6% of trials for ASD group; $t(103) = 1.8$, $p = .08$), or if RTs were less than 150 ms or greater than 2000 ms, were discarded. AI and EC effects were derived by subtracting the mean RT on compatible trials from that on incompatible trials.

Response time data. RTs on imitatively compatible trials [Control (mean ± standard error of the mean) = 493 ± 10 ms; ASD = 502 ± 9 ms] were faster than those on their respective incompatible trials [Control = 536 ±

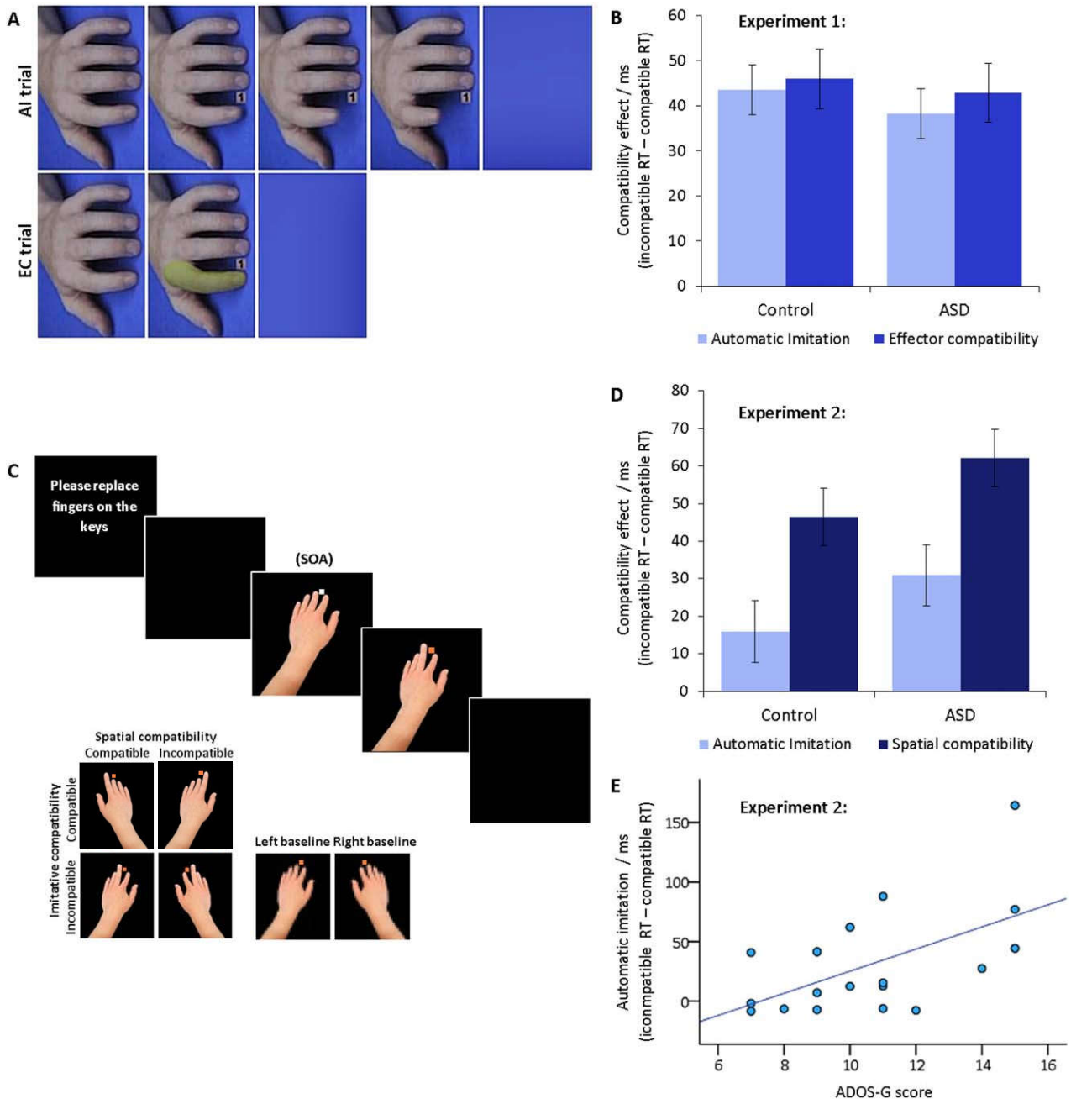

Figure 1. (A) Examples of automatic imitation (AI) and effector compatibility (EC) trials from Experiment 1 [Cook & Bird, 2012]. "1" indicates lift index finger, and thus the examples shown are from imitatively compatible and effector compatible trials. (B) AI and EC effects for both Control and ASD groups. (C) Examples of one trial structure from Experiment 2 and spatial and imitative compatibility of hand stimuli where an orange square indicates a required index finger lift. Movement hands presented at varying stimulus onset asynchronies (SOA: 1600, 2000, or 2400 ms) after the static hand. Hand stimuli from left and right baseline trials are also shown, during which the static hand is replaced by a pixelated hand eliciting no finger movement. (D) Automatic imitation and spatial compatibility effects for both Control and ASD groups. (E) Scatterplot of the correlation between autism symptom severity (ADOS-G scores) and AI from Experiment 2. NB. Error bars represent standard error of the mean.

12 ms; ASD = 541 ± 9 ms] for both control individuals [$t(44) = 6.4$, $P < 0.001$, $d = 1.0$] and individuals with ASD [$t(59) = 9.2$, $P < 0.001$, $d = 1.2$]. In addition, trials in which the finger highlighted by the green mask was compatible (EC compatible trials) with the required response [Control = 504 ± 9 ms; ASD = 514 ± 9 ms], were faster than their respective EC incompatible trials [Control = 550 ± 13 ms; ASD = 557 ± 10 ms] for both control

individuals [$t(44) = 5.9$, $P < 0.001$, $d = 0.9$] and individuals with ASD [$t(59) = 8.0$, $P < 0.001$, $d = 1.0$].

To assess any differences in these effects between groups, a two-way, mixed-model ANOVA was performed on the RT data for both AI (Group × AI) and EC (Group × EC) trials, with Group constituting the between subjects factor (Control, ASD) and AI and EC (compatible, incompatible) the respective within-subjects factors. These data confirmed significant main effects of AI [$F(1,103) = 115.1$, $P < 0.001$, $\eta_P^2 = 0.53$] and EC [$F(1,103) = 93.7$, $P < 0.001$, $\eta_P^2 = 0.48$]. However, there was no interaction between either AI and Group [$F(1,34) < 1$, $P = 0.49$, $\eta_P^2 = 0.002$] or EC and Group [$F(1,34) < 1$, $P = 0.74$, $\eta_P^2 = 0.001$]. Figure 1B shows AI and EC effects for both the Control and ASD groups.

Interim Discussion and Experiment 2

Experiment 1 successfully replicates previous findings of intact automatic imitation in a large sample of individuals diagnosed with ASD. However, despite the popularity of automatic imitation paradigms, some forms of these tests have been argued to be unsuitable to test imitation, due to the fact that the RT benefit observed on imitatively compatible trials may not be a product of automatic imitation but rather a product of spatial compatibility [Aicken, Wilson, Williams, & Mon-Williams, 2007; Bertenthal, Longo, & Kosobud, 2006; Jiménez et al., 2012]. Spatial compatibility effects elicited by the task-irrelevant spatial properties of task-relevant stimuli were first described by Simon and coworkers [Simon, 1968, 1969, Simon & Rudell, 1967] who showed that the spatial location of a stimulus (even if that location is task-irrelevant) facilitates movement responses on the same side of space. Thus, if the spatial location of an observed action matches the spatial location of the required response in an automatic imitation task, then it is possible that the faster responses to matching actions may be due to spatial compatibility rather than automatic imitation. While this criticism cannot explain some reports of intact automatic imitation in ASD [Bird et al., 2007], it may explain others [Hamilton et al., 2007; Press et al., 2010; Schunke et al., 2015; Spengler, Bird, et al., 2010], where group differences in automatic imitation may be masked by spatial compatibility.

The paradigm used in Experiment 1 and Cook and Bird [2011, 2012] rotates the stimulus hand with respect to the participant's hand to eliminate a simple spatial compatibility account of the observed automatic imitation effect. However, it is possible that results may be explained by orthogonal spatial compatibility effects whereby stimuli in the upper portion of space facilitate responses on the right side of space, and stimuli in the lower portion of space facilitate responses on the left side of space [Cho & Proctor, 2004].

Experiments with typical individuals have overcome these methodological problems by introducing a spatial compatibility control condition to demonstrate the presence of automatic imitation independently of simple and orthogonal spatial compatibility effects [Catmur & Heyes, 2011; Jiménez et al., 2012; Sowden & Catmur, 2015]. Experiment 2 utilizes one of these paradigms in order to re-examine automatic imitation in ASD. Typical, or even enhanced, imitation shown by individuals with ASD in this experiment cannot be due to spatial compatibility.

Method

Participants. Eighteen high-functioning individuals with a clinical diagnosis of ASD (2 female; 1 left-handed; $M_{AGE} = 37.7$ years, SD = 12.6) and an age, gender and IQ-matched sample of 18 healthy control individuals (3 female; 1 left-handed; $M_{AGE} = 34.9$ years, SD = 15.2) were recruited from a database held at the Institute of Psychiatry, Psychology, and Neuroscience, King's College London. All participants reported normal or corrected-to-normal vision. Two further individuals with ASD were excluded prior to data analysis as they made errors on more than 15% of trials [Sowden & Catmur, 2015], but all received a small monetary reward for their participation. Participants with ASD had previously received independent clinical diagnoses [according to DSM-IV criteria; American Psychiatric Association, 1994], and diagnosis was confirmed using the ADOS. AQ scores were significantly higher [$t(34) = 7.0$, $P < 0.001$] in the ASD group ($M = 35.1$, SD = 9.0) than the Control group ($M = 17.1$, SD = 6.3), and the two groups did not differ in age [$t(34) = 0.6$, $P = 0.55$], proportion of females [$\chi^2(1) = 0.23$, $P = 0.63$], or IQ [$t(34) = 0.3$, $P = 0.80$]. Full details of ASD and Control group characteristics are presented in Table 1.

Stimuli and Procedure. Experimental stimuli were the same as those used by Sowden and Catmur [2015] and presented in color on a black background (See Fig. 1C for full trial illustration). Task-irrelevant stimuli were images of a human right or left hand subtending a visual angle of 6.6° horizontally and either 8.5° (static and pixelated control hand), 9.1° (middle finger lifted) or 9.3° (index finger lifted) vertically. Middle and index finger lifts were superimposed onto the resting hand, subtending an angle of 0.6° and 0.7°, respectively and left hand stimuli were a direct mirror along the vertical plane of right hand stimuli.

The instantaneous presentation of the movement hand after the static hand produced apparent motion of the finger (lifting of the index or middle finger);

previously shown as a robust means of eliciting compatibility effects [Press et al., 2005]. Spatial compatibility (SC) was manipulated by using left and right hand stimuli. For example, in a spatially incompatible trial, a participant prompted to lift their right index finger would observe a left hand lifting its index finger. This allowed the spatial location of the observed finger movement to be manipulated independently from finger identity. Task-relevant (discriminative) stimuli comprised squares (occupying a 0.2° visual angle) colored orange or purple. A white square of identical dimensions operated as a fixation point (positioned halfway between the index and middle fingers of the static hand). The allocation of response cues (colored squares) to response options (index or middle finger lift) was counterbalanced across participants, with purple and orange squares indicating whether the participant should lift their index finger (from the "N" key) or middle finger (from the "M" key) on each trial.

On baseline trials the static hand was replaced by a pixelated hand which was designed so that it did not elicit SC or AI effects, but matched the transient, alerting visual change of the task-irrelevant movement hand in the standard trials [Wiggett, Downing, & Tipper, 2013].

Participants sat approximately 80cm from the laptop screen, placing their right arm (in the same orientation as the hand stimuli) on the table in front of them, and responses were made with the right hand using an external keyboard. Participants completed 10 practice trials, and were required to repeat these until achieving at least 80% accuracy. The main task consisted of 3 blocks of 36 trials, with each block lasting approximately 4 mins.

Hand stimuli were formally task irrelevant but allowed the independent manipulation of SC and AI. Thus, the on-screen hand performed either imitatively compatible or incompatible actions (AI manipulation) on the same or different side of space (SC manipulation) to the response required by the participant. Hand stimuli in the standard trials were manipulated in a 2 × 2 (AI × SC) design, resulting in four main trial types with a further two baseline trial types for left and right hand stimuli (see Fig. 1C). Each of these six trial types were presented 18 times in a randomized order across each block. A fully factorial combination of the six trial types, stimulus onset asynchronies and square color amounted to a total of 36 trials, which made up one full block in the experiment.

Results

Data were processed as in Sowden and Catmur [2015], with trials in which an inaccurate finger-lift response was made (4% of trials for Control group; 5% of trials for ASD group; $t(34) = 0.9$, $p = .40$), or if RT significantly deviated (±2.5 SD) from each participant's mean RT (<2% of trials), discarded. Mean RT and number of errors were calculated for each of the six trial types for control and ASD individuals. Compatibility effect data was derived by subtracting the mean RT on compatible trials from that on incompatible trials.

Response time data. RTs on imitatively compatible trials were faster than those on their respective incompatible trials for both the Control group [compatible = 517 ± 15 ms; incompatible = 533 ±19 ms; $t(17) = 2.7$, $P = 0.015$, $d = 0.6$] and the ASD group [compatible = 606 ± 39 ms; incompatible = 637 ± 48 ms; $t(17) = 2.9$, $P = 0.010$, $d = 0.7$]. Similarly, RTs on spatially compatible trials were faster than those on their respective incompatible trials for both the Control group [compatible = 502 ± 17 ms; incompatible = 548 ± 18 ms; $t(17) = 7.3$, $P < 0.001$, $d = 1.7$] and the ASD group [compatible = 591 ± 44 ms; incompatible = 653 ± 43 ms; $t(17) = 7.0$, $P < 0.001$, $d = 1.7$]. Separate two-way, repeated-measures ANOVAs were performed on the RT data for both groups, with within-subjects factors of SC (compatible, incompatible) and AI (compatible, incompatible). These analyses revealed a significant main effect of SC for both the Control group, $F(1,17) = 53.2$, $P < 0.001$, $\eta_P^2 = 0.76$, and the ASD group, $F(1,17) = 49.0$, $P < 0.001$, $\eta_P^2 = 0.74$, as well as a significant main effect of AI for both the Control, $F(1,17) = 7.4$, $P = 0.015$, $\eta_P^2 = 0.30$, and the ASD, $F(1,17) = 8.5$, $P = 0.010$, $\eta_P^2 = 0.33$, groups. There was no significant interaction between SC and AI for either the Control, $F(1,17) < 1$, $P = 0.74$, $\eta_P^2 = 0.007$, or the ASD, $F(1,17) < 1$, $P = 0.79$, $\eta_P^2 = 0.004$, groups. Figure 1D shows SC and AI effects (incompatible RTs – compatible RTs) for both the Control and ASD groups. Both effects are apparent on RTs, with SC effects typically larger than AI effects in both groups. SC and AI effects on RTs are of magnitudes consistent with previous studies using a similar paradigm [Catmur & Heyes, 2011; Sowden & Catmur, 2015].

To assess any differences in these effects between groups, a further three-way, mixed ANOVA (Group × SC × AI) was performed on the RT data, with Group constituting the between subjects factor (Control, ASD). These data once again confirmed significant main effects of SC, $F(1,34) = 98.8$, $P < 0.001$, $\eta_P^2 = 0.74$, and AI, $F(1,34) = 14.9$, $P < 0.001$, $\eta_P^2 = 0.30$, as well as no interaction between SC and AI, $F(1,34) < 1$, $P = 0.70$, $\eta_P^2 = 0.004$. Despite SC and AI effects being numerically larger in individuals with ASD, neither the main effect of Group, nor the interaction between Group and SC, $F(1,34) = 2.1$, $P = 0.16$, $\eta_P^2 = 0.06$, or Group and AI, $F(1,34) = 1.6$, $P = 0.22$, $\eta_P^2 = 0.04$, were significant. Finally,

no significant bivariate correlation was observed between SC and AI ($r = -0.04$, $P = 0.80$).

Hand × response effects. To ensure that trials did not elicit different RTs or response errors depending on whether a left or right task-irrelevant hand was presented or whether an index or middle finger response was required, a further two-way, repeated-measures ANOVA (hand × response) was carried out on the RT data from baseline trials. The within-subject factors were the hand presented (left, right) and the response required (index, middle). This analysis revealed no significant main effects of hand presented or response required, as well as no interaction between the two factors (all $P > 0.05$). These results suggest that the observed compatibility effects were unlikely to have been generated by general stimulus or response factors.

Autism severity and imitation effects. Although the AI effect observed in the RT data from the ASD group was not significantly larger than that observed in the control group (perhaps due to insufficient power to detect group differences), the AI effect was numerically larger in the ASD group than in the Control group. Thus, the relationship between autism severity and AI was assessed. ADOS severity scores showed a significant correlation with AI ($r = 0.56$, $P = 0.02$), with higher ASD severity scores correlating with higher AI effects. No correlation was observed between ADOS scores and SC ($r = -0.06$, $P = 0.82$), with a significant difference observed between these two correlation coefficients using Fisher's *r*-to-*z* test (*z*-score = 1.90, $P = 0.03$).

General Discussion

It has previously been argued that imitative behavior and MNS function are reduced in individuals with ASD [Williams et al., 2006]. However, previous investigations of these claims have often tested the performance of individuals with ASD on tests of voluntary imitation, which in addition to imitation require many abilities which have been shown to be impaired in ASD. Experiment 1 utilized a test of automatic imitation in a much larger sample of adults with ASD than obtained previously, and revealed automatic imitation to be intact in individuals with ASD and comparable to that demonstrated by a typical Control group. Due to the possibility that performance on the task used in Experiment 1 may be affected by orthogonal spatial compatibility, potentially masking any group differences in automatic imitation, Experiment 2 used a novel task allowing the independent manipulation and measurement of both automatic imitation and spatial compatibility. This not only allowed the observation that AI effects persist regardless of SC in both individuals with ASD and typical control individuals, but that AI (and not SC) effects are increased with increasing autism symptom severity, as indicated by ADOS scores.

Accordingly, our data demonstrate, contrary to the prevailing view [Williams, Whiten, Suddendorf, & Perrett, 2001, Williams et al., 2004], that individuals with autism show intact imitative behavior. In fact, enhanced AI (but not SC) effects were observed in association with higher autism severity. This finding is in line with symptoms of echolalia and echopraxia observed in many individuals with severe ASD, and with previous investigations showing *hyper* imitative performance in ASD [Bird et al., 2007; Grossi et al., 2012; Spengler, Bird, et al., 2010]. As it has been shown that AI effects as described here rely on the MNS [Catmur et al., 2009; Heiser et al., 2003; Heyes, 2011], and that performance on such experimental tasks of AI is associated with mimicry during everyday social interaction [Hogeveen & Obhi, 2012], these data speak to the longstanding debate regarding the integrity of the MNS in ASD. They suggest that the Broken Mirror Theory of autism is unlikely to be a valid explanation of the broad deficits observed in ASD. With the numerous methodological issues associated with neuroscientific investigations of MNS function in humans [Hamilton 2013; Coll et al., 2015], the present study provides some of the most promising evidence to date for intact MNS functioning in ASD.

Rather than characterizing ASD as due to a broken MNS, a more promising approach, in line with our findings of enhanced imitative behavior with increased autism severity, involves conceptualizing neurodevelopmental disorders like ASD as disorders of top-down control or modulation of social behavior [Bird, Catmur, Silani, Frith, & Frith, 2006; Cook, Barbalat, & Blakemore, 2012; Frith, 2003; Schunke et al., 2015; Sowden & Shah, 2014; Wang & Hamilton, 2012]. In accord with this notion, Cook and Bird [2012] demonstrated that the modulatory effects of pro-social priming on automatic imitation in typical individuals were absent in those with ASD. Similarly, Spengler, Bird, et al. [2010] found hyper-imitation to be correlated with impaired theory of mind in these individuals. Hence, a deficiency within a neural network which supports the top-down control of representations of "self" and "other" seems plausible in ASD and would account for hyperimitation as well as impairments in other socio-cognitive functions in these individuals [Sowden & Shah, 2014]. One such neural network suggested to subserve this role in the human brain involves the medial prefrontal cortex and temporoparietal junction [Brass et al., 2005; Santiesteban, Banissy, Catmur, & Bird, 2012; Spengler, von Cramon, & Brass, 2010]; a network whose role and function in self-other processing, and its interplay with the MNS, deserves attention in future autism research.

Conclusion

Imitation, and its neural substrate the MNS, has garnered a great deal of attention and controversy in the study of ASD. In contrast to the Broken Mirror Theory of ASD, this behavioral study demonstrated that automatic imitation is intact or even enhanced in individuals with ASD; indicative of a functional MNS. Experiment 1 established robust automatic imitation in a large sample of individuals with ASD. Experiment 2 isolated automatic imitation from that of spatial compatibility, to show that AI persists in both typical individuals and those with ASD when spatial compatibility is controlled for, and increases with autism symptom severity in autistic individuals. Thus, among an abundance of mixed neuroscientific evidence, AI tasks provide a pure and robust measure of imitative performance and offer some of the most promising evidence to date against a broken MNS in ASD.

Acknowledgments

The authors are extremely grateful to all the adults who took part in the studies, and to Punit Shah for his assistance with the testing in Experiment 2. Sophie Sowden was supported by a doctoral studentship from the Medical Research Council and Svenja Koehne was supported by a doctoral studentship from the German Research Foundation, DFG, grant number EXC302.

References

Aicken, M.D., Wilson, A.D., Williams, J.H., & Mon-Williams, M. (2007). Methodological issues in measures of imitative reaction times. Brain and Cognition, 63, 304–308.

American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders (3th ed.). Washington, DC: American Psychiatric Association Press.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Arlington, VA: American Psychiatric Publishing.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. Journal of Autism and Developmental Disorders, 31, 5–17.

Bertenthal, B.I., Longo, M.R., & Kosobud, A. (2006). Imitative response tendencies following observation of intransitive actions. Journal of Experimental Psychology: Human Perception and Performance, 32, 210–225.

Bird, G., Catmur, C., Silani, G., Frith, C., & Frith, U. (2006). Attention does not modulate neural responses to social stimuli in autism spectrum disorders. Neuroimage, 31, 1614–1624.

Bird, G., Leighton, J., Press, C., & Heyes, C. (2007). Intact automatic imitation of human and robot actions in autism spectrum disorders. Proceedings of the Royal Society B: Biological Sciences, 274, 3027–3031.

Brass, M., Bekkering, H., Wohlschläger, A., & Prinz, W. (2000). Compatibility between observed and executed finger movements: comparing symbolic, spatial, and imitative cues. Brain and Cognition, 44, 124–143.

Brass, M., Derrfuss, J., & von Cramon, D.Y. (2005). The inhibition of imitative and overlearned responses: A functional double dissociation. Neuropsychologia, 43, 89–98.

Brass, M., Ruby, P., & Spengler, S. (2009). Inhibition of imitative behavior and social cognition. Philosophical Transactions of the Royal Society B: Biological Sciences, 364, 2359–2367.

Catmur, C., & Heyes, C. (2011). Time course analyses confirm independence of imitative and spatial compatibility. Journal of Experimental Psychology: Human Perception and Performance, 37, 409–421.

Catmur, C., Walsh, V., & Heyes, C. (2009). Associative sequence learning: the role of experience in the development of imitation and the mirror system. Philosophical Transactions of the Royal Society B: Biological Sciences, 364, 2369–2380.

Cho, Y.S., & Proctor, R.W. (2004). Influences of multiple spatial stimulus and response codes on orthogonal stimulus—response compatibility. Attention, Perception, & Psychophysics, 66, 1003–1017.

Coll, M.P., Bird, G., Catmur, C., & Press, C. (2015). Cross-modal repetition effects in the mu rhythm indicate tactile mirroring during action observation. Cortex, 63, 121–131.

Cook, J., Barbalat, G., & Blakemore, S.J. (2012). Top-down modulation of the perception of other people in schizophrenia and autism. Frontiers in Human Neuroscience, 6, 175.

Cook, J., & Bird, G. (2011). Social attitudes differentially modulate imitation in adolescents and adults. Experimental Brain Research, 211, 601–612.

Cook, J.L., & Bird, G. (2012). Atypical social modulation of imitation in autism spectrum conditions. Journal of Autism and Developmental Disorders, 42, 1045–1051.

Cook, R., Bird, G., Catmur, C., Press, C., & Heyes, C. (2014). Mirror neurons: From origin to function. Behavioral and Brain Sciences, 37, 177–192.

Courchesne, E., Townsend, J., Akshoomoff, N.A., Saitoh, O., Yeung-Courchesne, R., Lincoln, A.J.,... & Lau, L. (1994). Impairment in shifting attention in autistic and cerebellar patients. Behavioral Neuroscience, 108, 848.

Frith, C. (2003). What do imaging studies tell us about the neural basis of autism? In G. Bock, & J. Goode (Eds.), Autism: Neural basis and treatment possibilities (pp. 149–166). Chichester, UK: Wiley.

Gowen, E., Bradshaw, C., Galpin, A., Lawrence, A., & Poliakoff, E. (2010). Exploring visuomotor priming following biological and non-biological stimuli. Brain and Cognition, 74, 288–297.

Grossi, D., Marcone, R., Cinquegrana, T., & Gallucci, M. (2012). On the differential nature of induced and incidental echolalia in autism. Journal of Intellectual Disability Research, 57, 903–912.

Hamilton, A.F.D.C. (2013). Reflecting on the mirror neuron system in autism: a systematic review of current theories. Developmental Cognitive Neuroscience, 3, 91–105.

Hamilton, A.F.D.C., Brindley, R.M., & Frith, U. (2007). Imitation and action understanding in autistic spectrum disorders: how valid is the hypothesis of a deficit in the mirror neuron system? Neuropsychologia, 45, 1859–1868.

Happé, F., & Frith, U. (1995). Theory of mind in autism. In Learning and cognition in autism (pp. 177–197). New York, NY: Springer.
Happé, F., Ronald, A., & Plomin, R. (2006). Time to give up on a single explanation for autism. Nature Neuroscience, 9, 1218–1220.
Hayes, K.J., & Hayes, C. (1952). Imitation in a home-reared chimpanzee. Journal of Comparative Psychology, 45, 450–459.
Heiser, M., Iacoboni, M., Maeda, F., Marcus, J., & Mazziotta, J.C. (2003). The essential role of Broca's area in imitation. European Journal of Neuroscience, 17, 1123–1128.
Heyes, C. (2011). Automatic imitation. Psychological Bulletin, 137, 463–483.
Hogeveen, J., & Obhi, S. S. (2012). Social interaction enhances motor resonance for observed human actions. Journal of Neuroscience, 32, 5984–5989.
Iacoboni, M., Woods, R.P., Brass, M., Bekkering, H., Mazziotta, J.C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. Science, 286, 2526–2528.
Jiménez, L., Recio, S., Méndez, A., Lorda, M.J., Permuy, B., & Méndez, C. (2012). Automatic imitation and spatial compatibility in a key-pressing task. Acta Psychologica, 141, 96–103.
Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: cognitive basis for stimulus-response compatibility – a model and taxonomy. Psychological Review, 97, 253–270.
Leighton, J., Bird, G., Charman, T., & Heyes, C. (2008). Weak imitative performance is not due to a functional 'mirroring' deficit in adults with Autism Spectrum Disorders. Neuropsychologia, 46, 1041–1049.
Leighton, J., & Heyes, C. (2010). Hand to mouth: automatic imitation across effector systems. Journal of Experimental Psychology: Human Perception and Performance, 36, 1174–1183.
Lord, C., Risi, S., Lambrecht, L., Cook Jr, E.H., Leventhal, B.L., DiLavore, P.C.,... & Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. Journal of Autism and Developmental Disorders, 30, 205–223.
McIntosh, D.N., Reichmann-Decker, A., Winkielman, P., & Wilbarger, J.L. (2006). When the social mirror breaks: deficits in automatic, but not voluntary, mimicry of emotional facial expressions in autism. Developmental Science, 9, 295–302.
Press, C., Bird, G., Flach, R.D., & Heyes, C. (2005). Robotic movement elicits automatic imitation. Cognitive Brain Research, 25, 632–640.
Press, C., Richardson, D., & Bird, G. (2010). Intact imitation of emotional facial actions in autism spectrum conditions. Neuropsychologia, 48, 3291–3297.
Prinz, W. (1997). Perception and action planning. European Journal of Cognitive Psychology, 9, 129–154.
Ramachandran, V.S., & Oberman, L.M. (2006). Broken mirrors: A theory of autism. Scientific American, 295, 62–69.
Rogers, S.J., & Pennington, B.F. (1991). A theoretical approach to the deficits in infantile autism. Development and Psychopathology, 3, 137–162.
Russell, J. (1997). Autism as an executive disorder. New York, NY: Oxford University Press.
Santiesteban, I., Banissy, M.J., Catmur, C., & Bird, G. (2012). Enhancing social ability by stimulating right temporoparietal junction. Current Biology, 22, 2274–2277.
Schunke, O., Schöttle, D., Vettorazzi, E., Brandt, V., Kahl, U., Bäumer, T.,... & Münchau, A. (2015). Mirror me: Imitative responses in adults with autism. Autism [Epub ahead of print]. doi: 10.1177/1362361315571757
Sebanz, N., Knoblich, G., Stumpf, L., & Prinz, W. (2005). Far from action-blind: Representation of others' actions in individuals with autism. Cognitive Neuropsychology, 22, 433–454.
Simon, J.R. (1968). Effect of ear stimulated on reaction time and movement time. Journal of Experimental Psychology, 78, 344–346.
Simon, J.R. (1969). Reactions towards the source of stimulation. Journal of Experimental Psychology, 81, 174–176.
Simon, J.R., & Rudell, A.P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. Journal of Applied Psychology, 51, 300–304.
Southgate, V., & Hamilton, A.F.D.C. (2008). Unbroken mirrors: Challenging a theory of autism. Trends in Cognitive Sciences, 12, 225–229.
Sowden, S., & Catmur, C. (2015). The role of the right temporoparietal junction in the control of imitation. Cerebral Cortex, 25, 1107–1113.
Sowden, S., & Shah, P. (2014). Self-other control: a candidate mechanism for social cognitive function. Frontiers in Human Neuroscience, 8, 789.
Spengler, S., Bird, G., & Brass, M. (2010). Hyperimitation of actions is related to reduced understanding of others' minds in autism spectrum conditions. Biological Psychiatry, 68, 1148–1155.
Spengler, S., von Cramon, D.Y., & Brass, M. (2010). Resisting motor mimicry: control of imitation involves processes central to social cognition in patients with frontal and temporo-parietal lesions. Social Neuroscience, 5, 401–416.
Stürmer, B., Aschersleben, G., & Prinz, W. (2000). Correspondence effects with manual gestures and postures: a study of imitation. Journal of Experimental Psychology: Human Perception and Performance, 26, 1746–1759.
Wang, Y., & Hamilton, A.F.D.C. (2012). Social top-down response modulation (STORM): a model of the control of mimicry in social interaction. Frontiers in Human Neuroscience, 6, 153.
Wiggett, A.J., Downing, P.E., & Tipper, S.P. (2013). Facilitation and interference in spatial and body reference frames. Experimental Brain Research, 225, 119–131.
Williams, D.L., Goldstein, G., Carpenter, P.A., & Minshew, N.J. (2005). Verbal and spatial working memory in autism. Journal of Autism and Developmental Disorders, 35, 747–756.
Williams, J.H., Waiter, G.D., Gilchrist, A., Perrett, D.I., Murray, A.D., & Whiten, A. (2006). Neural mechanisms of imitation and 'mirror neuron' functioning in autistic spectrum disorder. Neuropsychologia, 44, 610–621.
Williams, J.H., Whiten, A., & Singh, T. (2004). A systematic review of action imitation in autistic spectrum disorder. Journal of Autism and Developmental Disorders, 34, 285–299.
Williams, J.H., Whiten, A., Suddendorf, T., & Perrett, D.I. (2001). Imitation, mirror neurons and autism. Neuroscience & Biobehavioral Reviews, 25, 287–295.

2.1. Supplementary Data Analysis for Chapter 2

Analyses which were not included in the published article due to space limitations are presented here. In Experiment 1, no significant correlations were observed between automatic imitation and either ADOS (r = .100, p = .469) or AQ (r = -.087, p = .376) scores, and no significant correlation was observed between effector compatibility and either ADOS (r = -.048, p = .729) or AQ (r = -.038, p = .699) scores.

In Experiment 2, no significant correlation was observed between AQ scores and either imitative compatibility (r = -.026, p = .878) or spatial compatibility (r = .307, p = .068). Finally, no correlation was observed between the two measures of autistic trait severity (ADOS and AQ scores) in either Experiment 1 (r = -.064, p = .643) or Experiment 2 (r = -.265, p = .287).

Chapter 3. The specificity of the link between alexithymia, interoception, and imitation

This chapter is presented as a published article and is an exact copy of the following journal publication:

Sowden, S., Brewer, R., Catmur, C., & Bird, G. (2016). The specificity of the link between alexithymia, interoception, and imitation. *Journal of Experimental Psychology: Human Perception and Performance*, *42*, 1687-1692.

Corresponding Author:

Sophie Sowden, MRC Social, Genetic and Developmental Psychiatry, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, SE5 8AF, United Kingdom.

sophie.sowden@kcl.ac.uk

Journal of Experimental Psychology:
Human Perception and Performance
2016, Vol. 42, No. 11, 1687–1692

0096-1523/16/$12.00 http://dx.doi.org/10.1037/xhp0000310

OBSERVATION

The Specificity of the Link Between Alexithymia, Interoception, and Imitation

Sophie Sowden
King's College London

Rebecca Brewer
University of East London

Caroline Catmur
King's College London

Geoffrey Bird
King's College London and University College London

Alexithymia is a subclinical condition traditionally characterized by difficulties identifying and describing one's own emotions. Recent formulations of alexithymia, however, suggest that the condition may result from a generalized impairment in the perception of all bodily signals ("interoception"). Interoceptive accuracy has been associated with a variety of deficits in social cognition, but recently with an improved ability to inhibit the automatic tendency to imitate the actions of others. The current study tested the consequences for social cognition of the hypothesized association between alexithymia and impaired interoception by examining the relationship between alexithymia and the ability to inhibit imitation. If alexithymia is best characterized as a general interoceptive impairment, then one would predict that alexithymia would have the same relationship with the ability to control imitation as does interoceptive accuracy. Forty-three healthy adults completed measures of alexithymia, imitation-inhibition, and as a control, inhibition of nonimitative spatial compatibility. Results revealed the predicted relationship, such that increasing alexithymia was associated with an improved ability to inhibit imitation, and that this relationship was specific to imitation-inhibition. These results support the characterization of alexithymia as a general interoceptive impairment and shed light on the social ability of alexithymic individuals—with implications for the multitude of psychiatric, neurological, and neurodevelopmental disorders associated with high rates of alexithymia.

Keywords: alexithymia, interoception, imitation-inhibition, self–other processing

The ability to distinguish and control representations of the "self" and of "others" is important in almost all social contexts. For example, in order to avoid compulsive imitation of the actions of others one must accurately distinguish between other-related motor programs and those belonging to the self, and then control those representations such that representation of one's own motor program is enhanced and the other's inhibited. Similarly, when attempting to represent another's false belief, one must distinguish the mental state of the other from one's own, and then inhibit representation of one's own mental state and enhance that of the other (de Guzman, Bird, Banissy, & Catmur, 2016; Santiesteban et al., 2012; Sowden & Shah, 2014).

The cognitive processes supporting the control of self- and other-related representations are largely unknown, although it is possible that common processes are recruited across motor, cognitive, and affective domains (Spengler, Bird, & Brass, 2010). Causal evidence for such a link was provided by Santiesteban and colleagues (2012), who demonstrated that training to inhibit imitation resulted in an improved ability to take the visual perspective of another, that is,

Sophie Sowden, MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London; Rebecca Brewer, School of Psychology, University of East London; Caroline Catmur, Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London; Geoffrey Bird, MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, and Institute of Cognitive Neuroscience, University College London.

We are extremely grateful to all the adults who took part in our study, and we would like to thank Karl Friston for helpful discussions regarding predictive coding models.

Sophie Sowden is supported by a doctoral studentship from the Medical Research Council, which provided the funds for participant payments in this study. Geoffrey Bird is supported by the Baily Thomas Charitable Trust.

Correspondence concerning this article should be addressed to Sophie Sowden, Institute of Psychiatry, Psychology and Neuroscience, King's College London, 16 De Crespigny Park, London SE5 8AF, United Kingdom. E-mail: sophie.sowden@kcl.ac.uk

participants were better able to separate their own and another's visual perspective, and selectively represent that of the other. Similar evidence was provided by de Guzman and colleagues, who demonstrated that training to inhibit imitation enhanced empathic responses to another's pain (de Guzman et al., 2016). A mechanism of self–other control has been suggested to explain results such as these, which enables selective representation of self- or other-relevant representations such that the self is enhanced and the other inhibited (e.g., in the case of inhibiting imitation), or the other enhanced and the self inhibited (in the case of perspective taking or empathy for pain), according to task demands (Hogeveen et al., 2015; Happé, Cook, & Bird, in press).

Recent computational models within the predictive coding framework suggest that the ability to distinguish between self and other is dependent upon interoception, the perception of the internal state of one's own body (Seth, 2013; Quattrocki & Friston, 2014). These theories suggest that interoceptive information is used to build representations which correspond to one's "sentient, feeling self" (Seth, 2013). Good awareness of interoceptive cues is therefore thought to be crucial for the awareness of one's own body, and the representation of oneself as distinct from others. Lower interoceptive accuracy would therefore be expected to result in a reduced ability to distinguish between self and other.

One study is notable, however, in that its results argue against such a relationship between interoceptive accuracy and self–other processing. Ainley, Brass, and Tsakiris (2014) investigated the relationship between interoceptive accuracy and the ability to inhibit imitation. As described previously, when inhibiting the automatic tendency to imitate, one must distinguish one's own motor intention from that of the other, and then enhance representation of the self and inhibit representation of the other (Brass, Derrfuss, & von Cramon, 2005; Brass, Ruby, & Spengler, 2009; Spengler, Brass, Kühn, & Schütz-Bosbach, 2010; Spengler et al., 2010). Rather than the predicted positive relationship between interoceptive accuracy and ability to inhibit imitation, Ainley et al. (2014) found a negative relationship such that those with superior interoceptive accuracy had greater difficulty inhibiting the tendency to imitate, suggesting poorer ability to distinguish between self and other.

It has recently been suggested that alexithymia, a subclinical condition characterized by an inability to identify and describe one's own emotions (Sifneos, 1973), is best characterized as a generalized impairment of interoception (Brewer, Happé, Cook, & Bird, 2015; Bird & Viding, 2014). This claim is supported by recent demonstrations of lower accuracy in detecting one's heartbeat in those with high levels of alexithymia (Herbert, Herbert, & Pollatos, 2011; Shah, Hall, Catmur, & Bird, 2016), along with reduced accuracy when reporting one's degree of physiological arousal (Gaigg, Cornell, & Bird, in press), as well as a failure to recognize nonaffective interoceptive states such as fatigue, temperature, hunger, and satiety (Brewer, Cook, & Bird, in press). If the hypothesis that alexithymia is characterized by reduced interoceptive accuracy is correct, one would therefore expect the impact of alexithymia to be consistent with that of interoceptive accuracy. In other words, if alexithymia is the result of generally lower interoceptive accuracy, then one would expect it to be associated with the same pattern of abilities and impairments that have previously been demonstrated to be associated with interoceptive accuracy. This is the focus of the current study.

The Current Study

The results obtained by Ainley et al. (2014) are of theoretical importance; if greater interoceptive accuracy is associated with reduced ability to inhibit imitation, then current theoretical models linking interoception to improved self–other distinction require revision. For the present purposes, however, they provide an opportunity to investigate the functional consequences of the proposed link between alexithymia and interoception. If high levels of alexithymia reflect reduced interoceptive accuracy, then one would expect increasing alexithymia to predict an increased ability to inhibit imitation, as found by Ainley et al. (2014) for interoception.

In addition, the current study is able to investigate the specificity of the link between alexithymia and the inhibition of imitation. Ainley et al. (2014) used a test of imitation in which participants are required to perform either an index- or middle-finger lift. Simultaneously with the participant's response, a hand presented on the computer screen performed the same action ("imitatively compatible trials") or a different action ("imitatively incompatible trials"). The reaction time (RT) difference between compatible and incompatible trials is taken as an index of the participant's ability to inhibit the automatic tendency to imitate the action of the computer hand. In the version of the task used by Ainley et al. (2014), however, imitatively compatible stimuli (e.g., an observed middle-finger lift when a middle-finger lift is required) were also spatially compatible with the required response. There is now ample evidence (Catmur & Heyes, 2011; Cho & Proctor, 2004; Sowden & Catmur, 2015) that stimuli presented on the same side of space as a required response, regardless of their identity, prompt faster responses than those on the opposite side of space. When using such a design, it is impossible to distinguish spatial from imitative compatibility. We therefore utilized a paradigm able to distinguish between imitative and spatial compatibility (Sowden & Catmur, 2015). If effects of interoception and, by hypothesis, alexithymia, are specific to imitation-inhibition, and not more generally to nonimitative stimulus–response compatibility, then we should find effects of alexithymia on imitative, but not spatial, compatibility.

Method

Ethics. The experiment was approved by the local research ethics committee and was performed in accordance with the principles of the Helsinki Declaration (World Medical Association, 2013). All participants provided written informed consent and were aware they could withdraw at any time.

Participants. Forty-three healthy adult participants (20 males, mean age = 29.6 years, standard deviation [*SD*] = 12.4) were recruited via King's College London research recruitment systems. All reported being right-handed, with normal or corrected-to-normal vision, and were typically staff or students affiliated with King's College London. One further participant was excluded prior to data analysis as that participant made more than 15% errors on the imitation-inhibition task (Sowden & Catmur, 2015).

Imitation-inhibition task. The stimuli used in the imitation-inhibition task were identical to those used by Sowden and Catmur (2015), presented in color on a black background (see Figure 1 for full trial and stimuli illustrations), on a 15.6-inch LCD laptop screen via E-Prime2 (Psychology Software Tools, Sharpsburg, PA). Images of human right and left hands served as the task-irrelevant stimuli (horizontal visual angle of 6.5°), consisting of static hands (where all fingers were at rest) and, on standard trials, hands for which the index

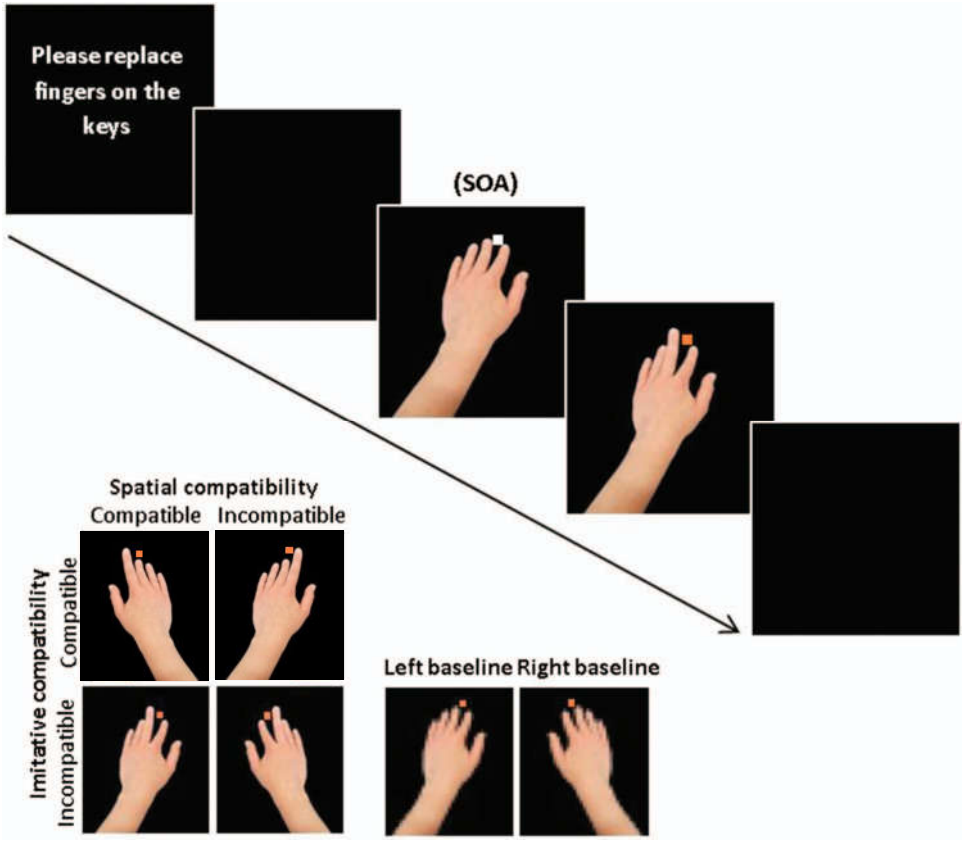

Figure 1. Example of one full trial in the experiment and of the task-irrelevant hand stimuli. Stimulus onset asynchronies (SOA) were 1,600, 2,000, or 2,400 ms. Labels denote spatial and imitative compatibility of stimuli on the standard trials, illustrating the 2 × 2 design, and left- or right-hand stimuli on the baseline trials, when the orange (dark) square indicates a required index-finger lift. When a middle-finger lift is required, levels of spatial and imitative compatibility are each reversed. For this response mapping, the trial illustrated is imitatively incompatible and spatially compatible, whereas for the response mapping for which orange indicates a middle-finger lift, it is imitatively compatible and spatially incompatible. See the online article for the color version of this figure.

or middle fingers were in a lifted position (subtending vertical visual angles of 9.4° and 9.2°, respectively). The immediate transition from a static hand to a finger-lift image produces apparent motion of the finger (Press, Bird, Flach, & Heyes, 2005). On baseline trials the static hand was replaced by a pixelated hand which was designed not to elicit spatial or imitative compatibility effects, but precisely matched the timing of the task-irrelevant movement in standard trials.

Task-relevant discriminative stimuli which indicated the participant's required response were orange or purple squares (occupying 0.2° visual angle). A white square of identical dimensions, positioned between the index and middle fingers of the static hand, functioned as a fixation point. Allocation of discriminative stimuli to response options (index- or middle-finger lift) was counterbalanced across participants, with purple and orange squares indicating whether the participant should lift their index finger (from the *N* key) or middle finger (from the *M* key) on each trial.

Participants sat approximately 80 cm from the screen, placing their right arm (in the same orientation as the hand stimuli) on the table in front of them, and responses were made with the right hand on an external keyboard. Participants completed 10 practice trials, and were required to repeat these until ≥80% accuracy was achieved. The main task consisted of three blocks of 36 trials, with each block lasting approximately 4 min.

Left- and right-hand stimuli were a direct mirror of one another along the vertical axis, and allowed the independent manipulation of spatial compatibility of the hand stimuli, that is, manipulation of the spatial location of the observed finger movement independent from finger identity. On each trial, the task-irrelevant hand performed either imitatively compatible or incompatible actions (imitative compatibility manipulation) on the same or different side of space (spatial compatibility manipulation) to the response required by the participant. Hand stimuli in the standard trials were manipulated in a 2 × 2 (imitatively compatibility × spatial compatibility) design. This produced four main trial types with a further two baseline trial types for left- and right-hand stimuli. In an imitatively compatible trial, a participant prompted to lift his or her right index finger may observe a right hand also lifting its index finger (this trial is also spatially compatible), or he or she may observe a left hand lifting its index finger (this trial is spatially incompatible). In an imitatively incompatible trial, the right-hand index finger response may be performed during the observation of a left hand lifting its middle finger, with this trial being spatially compatible with the required right-hand index

finger lifting response; or during the observation of a right hand lifting its middle finger, with this trial being spatially incompatible with the required response.

Questionnaire measures. Prior to completing the task, participants completed both the Toronto Alexithymia Scale (TAS-20; Bagby, Taylor, & Parker, 1994) and the Autism Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001). Typically, levels of alexithymia, as indexed by the TAS-20, and levels of autistic traits, as indexed by the AQ, are correlated (Aaron, Benson, & Park, 2015), and autistic traits have been predicted to be related to the inhibition of imitation by Quattrocki and Friston (2014). Thus, in the current study we incorporated both measures to identify the variance in performance accounted for by levels of alexithymia over and above that accounted for by autistic traits.

Results

Alexithymia. The mean sample TAS-20 score observed (46.42) was similar to population figures, but with a wider distribution (SD = 14.79; population M = 45.57, SD = 11.35; Parker, Taylor, & Bagby, 2003). AQ scores showed the same pattern (M = 18.35; SD = 8.74; population M = 16.94, SD = 5.59; Ruzich et al., 2015).

Imitation-inhibition task. Mean RT was calculated for each trial type. RTs for spatially and imitatively compatible trials were faster than their incompatible counterparts (Figure 2a), and consistent with RTs observed previously (Sowden & Catmur, 2015; Sowden et al., 2015). These were analyzed using a two-way, repeated-measures analysis of variance (ANOVA), with within-subject factors of spatial compatibility (compatible, incompatible) and imitative compatibility (compatible, incompatible), which revealed significant main effects of both spatial compatibility, $F(1, 42) = 136.9$, $p < .001$, $\eta_p^2 = .77$, and imitative compatibility, $F(1, 42) = 20.57$, $p < .001$, $\eta_p^2 = .33$. There was no significant interaction between spatial and imitative compatibility ($p = .845$).

Simple bivariate correlations revealed that alexithymia (as indicated by TAS-20 scores) was significantly associated with imita-

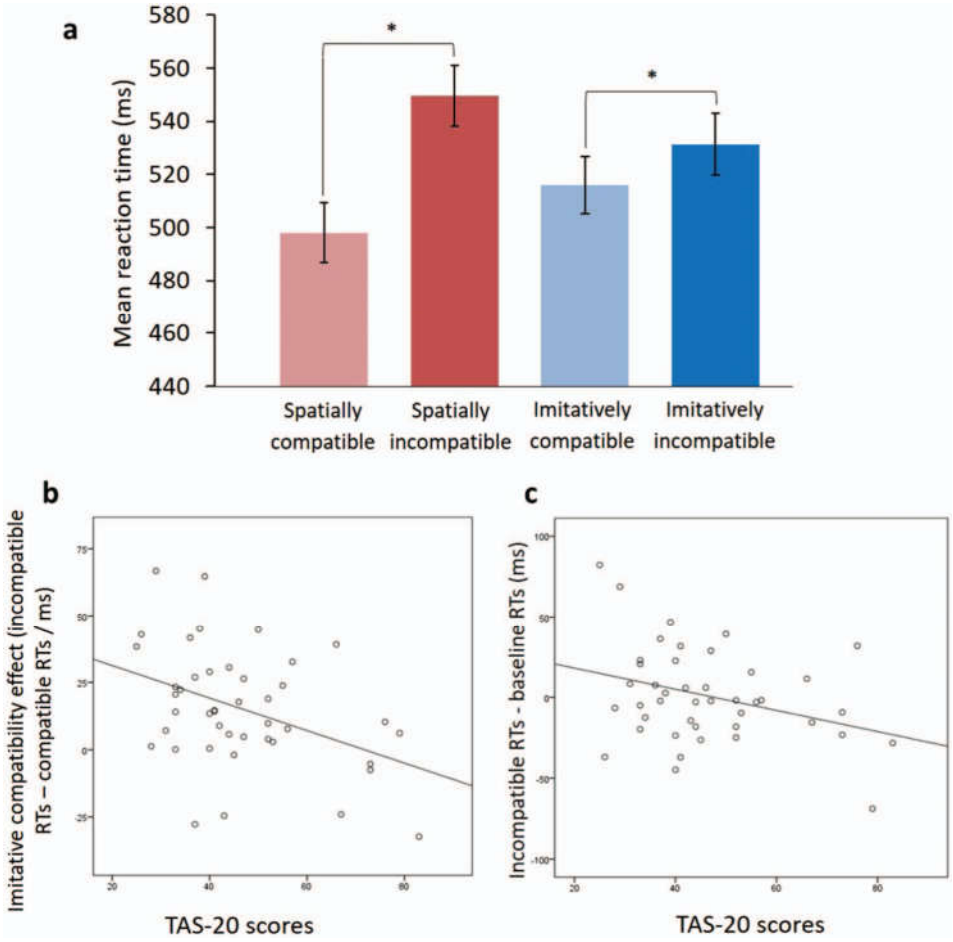

Figure 2. (a) Mean reaction times (RTs; in milliseconds) for each of the four main trial types (spatially compatible, spatially incompatible, imitatively compatible, and imitatively incompatible). Error bars represent the standard error of the mean. Significance at $p < 0.001$ is denoted by *. (b) The relationship between imitative compatibility effects (incompatible RTs—compatible RTs/ms) and Toronto Alexithymia Scale (TAS-20) alexithymia scores. (c) The relationship between the degree of response slowing on imitatively incompatible trials (imitatively incompatible RTs—baseline RTs/ms) and TAS-20 alexithymia scores. See the online article for the color version of this figure.

tive compatibility ($r = -.404$, $p = .007$), whereby as alexithymia scores increased, imitative compatibility decreased (Figure 2b). Spatial compatibility was not associated with alexithymia ($r = -.168$, $p = .281$). When imitatively compatible and incompatible trials were assessed separately (after subtracting RT on baseline trials), only the imitatively incompatible trials were significantly associated with alexithymia (imitatively incompatible trials, $r = -.325$, $p = .031$; imitatively compatible trials, $r = .028$, $p = .857$). As alexithymia scores increased, the degree to which observed incompatible actions interfered with participant responses decreased (Figure 2c). These results are very much in line with those found by Ainley and colleagues (Ainley et al., 2014)—who found a correlation coefficient for the association between interoceptive accuracy and imitation-inhibition of .41—and are thus consistent with the hypothesis that higher levels of alexithymia reflect decreased interoceptive accuracy.

Alexithymic and autistic traits were positively correlated ($r = .404$, $p = .007$), consistent with existing literature. Variables such as age, gender, and overall mean RT also influence performance on basic cognitive tasks (Harms et al., 2010). Thus, hierarchical regression analyses were conducted to test whether alexithymia accounts for variance over and above that explained by these other factors.

Two hierarchical regressions were conducted to model the variance in the size of the imitative compatibility effect and RT on imitatively incompatible trials, respectively. The demographic variables (age, gender, mean RT, and AQ scores) were entered into the first step of the regression models. Neither AQ nor the interaction between AQ and TAS scores were significant in their association with either the imitative compatibility effect or performance on imitatively incompatible trials ($ps > .05$) and first-level models for both regression analyses were not significant ($p = .249$ and $p = .207$, respectively). Alexithymia scores were entered into the second step of the model and revealed alexithymia to be a significant predictor both of the imitative compatibility effect, $\beta = -.650$, $t(42) = -2.81$, $p = .008$, and of RT on imitatively incompatible trials, $\beta = -.819$, $t(42) = -2.67$, $p = .011$. The addition of alexithymia scores significantly improved the fit of both models, increasing the variance accounted for by 18.5% for the imitative compatibility model, $F(1, 37) = 7.89$, $p = .008$, and by 18.2% for the incompatible trials model, $F(1, 37) = 7.12$, $p = .011$.

Discussion

The present study sought to investigate the relationship between alexithymia and the ability to inhibit imitation, based on an association between imitation-inhibition and interoceptive accuracy (Ainley et al., 2014) and the hypothesis that alexithymia is characterized by interoceptive impairment (Bird & Viding, 2014; Brewer et al., 2015). Moreover, the specificity of the link between alexithymia and imitation-inhibition was investigated through the use of a task in which imitative and spatial compatibility effects could be dissociated.

Results were as predicted by models suggesting that alexithymia is a product of general interoceptive deficits; increasing alexithymia was associated with improved ability to inhibit imitation in the same way, and to the same degree, as interoceptive accuracy (Ainley et al., 2014). The relationship between alexithymia and imitation-inhibition was specific to imitatively incompatible trials, suggesting that performance was driven by the ability to distinguish and control representations of one's own motor intention from that of the other, rather than a tendency to imitate. If the association was driven by imitation, rather than imitation inhibition, then one would also expect to see an association between alexithymia and RT on imitatively compatible trials.

Furthermore, the relationship between alexithymia and the inhibition of imitation was specific to inhibition of imitative responses. There was no relationship between alexithymia and spatial compatibility, which is particularly striking as responding on spatially incompatible trials necessitates inhibition of an automatic stimulus-response mapping—a task closely matched to that necessary on imitatively incompatible trials. This finding provides further confidence in the attribution of effects to self- and other-related processing, rather than general executive function or motor inhibition ability.

These findings contribute to our understanding of the relationship among alexithymia, interoception, and social ability. While supporting the link between alexithymia and reduced interoceptive accuracy, they suggest that alexithymia may be characterized by increased imitative control according to task or situational demands. Although imitation in social situations is generally considered to promote affiliation—for example, people who engage in imitative behavior are rated as more likable than those who do not (Chartrand & Bargh, 1999)—humans do not always imitate. Motivational and situational factors play an important role in modulating the extent to which people imitate others. For example, individuals are more likely to engage in mimicry with in-group members if they share a common goal and need to cooperate, but not when they are in competition (LaFrance, 1985; although see Marsh, Bird, & Catmur, 2016). Our results suggest that individuals with alexithymia may be better able to precisely modulate their degree of imitation in such situations, leading to a more selective impact on social relationships.

Conclusion

The present study supports the suggestion that interoceptive atypicalities observed in individuals with alexithymia may not be selective to interpreting one's own emotions, but may in fact be associated with more general interoceptive difficulties. This conclusion is of clinical interest, as impaired interoception may provide an explanation for the symptom commonalities seen across a large number of neurological, neurodevelopmental, and psychiatric disorders, all characterized by high rates of alexithymia. Given the link between alexithymia and interoceptive awareness (Herbert et al., 2011; Shah et al., 2016), our findings are consistent with those of Ainley and colleagues, supporting the link between interoceptive accuracy and the inhibition of imitation behavior, while providing an extension to this work by demonstrating the specificity of this link to self- and other-related processing rather than general executive function or motor inhibition. These findings are significant for our understanding of the specificity of interoceptive and behavioral profiles of individuals with alexithymia, but may also necessitate revision of current theoretical models of the relationship between interoception and sociocognitive ability.

References

Aaron, R. V., Benson, T. L., & Park, S. (2015). Investigating the role of alexithymia on the empathic deficits found in schizotypy and autism spectrum traits. *Personality and Individual Differences, 77,* 215–220. http://dx.doi.org/10.1016/j.paid.2014.12.032

Ainley, V., Brass, M., & Tsakiris, M. (2014). Heartfelt imitation: High interoceptive awareness is linked to greater automatic imitation. *Neuropsychologia, 60,* 21–28. http://dx.doi.org/10.1016/j.neuropsychologia.2014.05.010

Bagby, R. M., Taylor, G. J., & Parker, J. D. (1994). The twenty-item Toronto Alexithymia Scale—II. Convergent, discriminant, and concurrent validity. *Journal of Psychosomatic Research, 38,* 33–40. http://dx.doi.org/10.1016/0022-3999(94)90006-X

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31,* 5–17. http://dx.doi.org/10.1023/A:1005653411471

Bird, G., & Viding, E. (2014). The self to other model of empathy: Providing a new framework for understanding empathy impairments in psychopathy, autism, and alexithymia. *Neuroscience and Biobehavioral Reviews, 47,* 520–532. http://dx.doi.org/10.1016/j.neubiorev.2014.09.021

Brass, M., Derrfuss, J., & von Cramon, D. Y. (2005). The inhibition of imitative and overlearned responses: A functional double dissociation. *Neuropsychologia, 43,* 89–98. http://dx.doi.org/10.1016/j.neuropsychologia.2004.06.018

Brass, M., Ruby, P., & Spengler, S. (2009). Inhibition of imitative behaviour and social cognition. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 364,* 2359–2367. http://dx.doi.org/10.1098/rstb.2009.0066

Brewer, R., Cook, R., & Bird, G. (in press). *Alexithymia: A general deficit of interoception*. Royal Society Open Science.

Brewer, R., Happé, F., Cook, R., & Bird, G. (2015). Commentary on "Autism, oxytocin and interoception": Alexithymia, not autism spectrum disorders, is the consequence of interoceptive failure. *Neuroscience and Biobehavioral Reviews, 56,* 348–353. http://dx.doi.org/10.1016/j.neubiorev.2015.07.006

Catmur, C., & Heyes, C. (2011). Time course analyses confirm independence of imitative and spatial compatibility. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 409–421. http://dx.doi.org/10.1037/a0019325

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology, 76,* 893–910. http://dx.doi.org/10.1037/0022-3514.76.6.893

Cho, Y. S., & Proctor, R. W. (2004). Influences of multiple spatial stimulus and response codes on orthogonal stimulus-response compatibility. *Perception & Psychophysics, 66,* 1003–1017. http://dx.doi.org/10.3758/BF03194991

de Guzman, M., Bird, G., Banissy, M. J., & Catmur, C. (2016). Self-other control processes in social cognition: From imitation to empathy. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 371,* 20150079. http://dx.doi.org/10.1098/rstb.2015.0079

Gaigg, S. B., Cornell, A. S. F., & Bird, G. (in press). The psychological mechanisms of alexithymia in autism spectrum disorder. *Autism: An International Journal of Research and Practise*.

Happé, F., Cook, J., & Bird, G. (in press). The structure of social cognition: In(ter)dependence of sociocognitive processes. *Annual Review of Psychology*.

Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies. *Neuropsychology Review, 20,* 290–322. http://dx.doi.org/10.1007/s11065-010-9138-6

Herbert, B. M., Herbert, C., & Pollatos, O. (2011). On the relationship between interoceptive awareness and alexithymia: Is interoceptive awareness related to emotional awareness? *Journal of Personality, 79,* 1149–1175. http://dx.doi.org/10.1111/j.1467-6494.2011.00717.x

Hogeveen, J., Obhi, S. S., Banissy, M. J., Santiesteban, I., Press, C., Catmur, C., & Bird, G. (2015). Task-dependent and distinct roles of the temporoparietal junction and inferior frontal cortex in the control of imitation. *Social Cognitive and Affective Neuroscience, 10,* 1003–1009. http://dx.doi.org/10.1093/scan/nsu148

LaFrance, M. (1985). Postural mirroring and intergroup relations. *Personality and Social Psychology Bulletin, 11,* 207–217. http://dx.doi.org/10.1177/0146167285112008

Marsh, L. E., Bird, G., & Catmur, C. (2016). The imitation game: Effects of social cues on 'imitation' are domain-general in nature. *NeuroImage, 139,* 368–375. http://dx.doi.org/10.1016/j.neuroimage.2016.06.050

Parker, J. D., Taylor, G. J., & Bagby, R. M. (2003). The 20-Item Toronto Alexithymia Scale. III. Reliability and factorial validity in a community population. *Journal of Psychosomatic Research, 55,* 269–275. http://dx.doi.org/10.1016/S0022-3999(02)00578-0

Press, C., Bird, G., Flach, R., & Heyes, C. (2005). Robotic movement elicits automatic imitation. *Cognitive Brain Research, 25,* 632–640. http://dx.doi.org/10.1016/j.cogbrainres.2005.08.020

Quattrocki, E., & Friston, K. (2014). Autism, oxytocin and interoception. *Neuroscience and Biobehavioral Reviews, 47,* 410–430. http://dx.doi.org/10.1016/j.neubiorev.2014.09.012

Ruzich, E., Allison, C., Smith, P., Watson, P., Auyeung, B., Ring, H., & Baron-Cohen, S. (2015). Measuring autistic traits in the general population: A systematic review of the Autism-Spectrum Quotient (AQ) in a nonclinical population sample of 6,900 typical adult males and females. *Molecular Autism, 6,* 2.

Santiesteban, I., White, S., Cook, J., Gilbert, S. J., Heyes, C., & Bird, G. (2012). Training social cognition: From imitation to theory of mind. *Cognition, 122,* 228–235. http://dx.doi.org/10.1016/j.cognition.2011.11.004

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences, 17,* 565–573. http://dx.doi.org/10.1016/j.tics.2013.09.007

Shah, P., Hall, R., Catmur, C., & Bird, G. (2016). Alexithymia, not autism, is associated with impaired interoception. *Cortex, 81,* 215–220. http://dx.doi.org/10.1016/j.cortex.2016.03.021

Sifneos, P. E. (1973). The prevalence of 'alexithymic' characteristics in psychosomatic patients. *Psychotherapy and Psychosomatics, 22,* 255–262. http://dx.doi.org/10.1159/000286529

Sowden, S., & Catmur, C. (2015). The role of the right temporoparietal junction in the control of imitation. *Cerebral Cortex, 25,* 1107–1113. http://dx.doi.org/10.1093/cercor/bht306

Sowden, S., Koehne, S., Catmur, C., Dziobek, I., & Bird, G. (2016). Intact automatic imitation and typical spatial compatibility in autism spectrum disorder: Challenging the broken mirror theory. *Autism Research, 9,* 292–300. http://dx.doi.org/10.1002/aur.1511

Sowden, S., & Shah, P. (2014). Self-other control: A candidate mechanism for social cognitive function. *Frontiers in Human Neuroscience, 8,* 789. http://dx.doi.org/10.3389/fnhum.2014.00789

Spengler, S., Bird, G., & Brass, M. (2010). Hyperimitation of actions is related to reduced understanding of others' minds in autism spectrum conditions. *Biological Psychiatry, 68,* 1148–1155. http://dx.doi.org/10.1016/j.biopsych.2010.09.017

Spengler, S., Brass, M., Kühn, S., & Schütz-Bosbach, S. (2010). Minimizing motor mimicry by myself: Self-focus enhances online action-control mechanisms during motor contagion. *Consciousness and Cognition, 19,* 98–106. http://dx.doi.org/10.1016/j.concog.2009.12.014

World Medical Association. (2013). *WMA declaration of Helsinki: Ethical principles for medical research involving human subjects*. Retrieved from http://www.wma.net/en/30publications/10policies/b3/

Received May 4, 2016
Revision received August 11, 2016
Accepted August 12, 2016 ■

Chapter 4. Brief introduction to transcranial direct current stimulation and relevant methodological issues

Transcranial direct current stimulation (tDCS) is a non-invasive form of brain stimulation, which induces a constant weak electrical current in the cortex via two conductive rubber-coated electrodes, placed on the scalp inside saline-soaked sponges and connected to a battery-powered constant current stimulator. Such a weak electrical current is used to modulate neuronal cortical excitability and the resulting effects on behaviour and cognition are observed. This chapter provides an introduction to the history, physiology and methodological issues concerned with the use of tDCS.

4.1. A Brief History of tDCS

The use of electrical stimulation has its roots as early as 43 AD. The physician Scribonius Largus experimented with the use of live torpedo fish, which produce strong direct electrical currents, to alleviate headaches when placed on an individual's forehead (Debru, 2006; Kellaway, 1946; Priori, 2003). Methods advanced during the late 18th century with the development of the electric battery by Italian scientist Alessandro Volta, as well as his work with Luigi Galvani on the role of electricity and galvanic currents to induce muscle movement and maintain muscle physiology in animals (Parent, 2004). The development of a device to deliver a constant, direct electrical current prompted a body of work by Galvani's nephew Giovanni Aldini in applying galvanic currents to patients suffering from melancholia, symbolising the first clinical applications for direct current stimulation. Without means to test the intensity of direct currents, Aldini experimented on himself to establish parameters which he then reported to be successful in alleviating symptoms in patients (Priori, 2003). However, there was mixed success in this work as well as with the subsequent application of direct current stimulation in a number of other psychiatric disorders (Lolas, 1977). Thus, the use of

direct current stimulation fell out of the spotlight with the rising popularity of electroconvulsive therapy as well as the development of drugs to target clinical disorders between 1930 and 1960.

Direct current stimulation applied transcranially rose in popularity once again in the 1960s, when researchers began to recognise the benefit of understanding the mechanisms behind the effects of weak, direct electrical currents on the brain (Priori, 2003). For example, important discoveries concerning brain physiology were made by investigating the effect of anodal and cathodal tDCS on cortical excitability in animals (Albert, 1966; Creutzfeldt, Fromm, & Kapp, 1962; Purpura & McMurtry, 1965), as well as some early findings on the impact of tDCS on animal and human cognition (Lippold & Redfearn, 1964; Redfearn, Lippold, & Costain, 1964).

To complete a full circle of the history and popularity of electrical stimulation, much of our current understanding of the effects of tDCS has come from the increased use of tDCS as a neuroscientific tool for understanding brain and behaviour in the past two decades. Not only has it been studied extensively to aid understanding of human motor functions (Nitsche & Paulus, 2001; Priori, Berardelli, Rona, Accornero, & Manfredi, 1998), but also the neurocognitive mechanisms involved in various executive functions (Fregni et al., 2005; Gill, Shah-Basak, & Hamilton, 2015; Zaehle, Sandmann, Thorne, Jäncke, & Herrmann, 2011), language (Monti et al., 2013; Price, McAdams, Grossman, & Hamilton, 2015) and even aspects of social functioning (Sellaro, Nitsche, & Colzato, 2016, 2017). Finally, due to the non-invasive, painless nature of tDCS, there has been recent interest and success in the utility of tDCS to treat depression (Nitsche, Boggio, Fregni, & Pascual-Leone, 2009; Shiozawa et al., 2014) and some other clinical conditions such as stroke, Parkinson's disease and pain-related conditions (Borckardt et al., 2011, 2012; Demirtas-Tatlidede, Vahabzadeh-Hagh, & Pascual-Leone, 2013; Fregni et al., 2006; Lindenberg, Renga, Zhu, Nair, & Schlaug, 2010).

4.2. Physiological Basis of tDCS

TDCS impacts brain function, not by inducing neuronal action potentials as is the case with transcranial magnetic stimulation (TMS), but by hyperpolarising or depolarising resting membrane potentials of neurons in the underlying cortex. This in turn modifies the spontaneous firing of these neurons (Nitsche & Paulus, 2000). This is achieved by delivering a sub-threshold, weak electric current (between 1 and 2 mA) to the brain, for a period of approximately 20 minutes, using a battery-powered constant current stimulator connected to two surface rubber-conducting electrodes placed on the scalp inside saline-soaked sponges (DaSilva, Volz, Bikson, & Fregni, 2011). One electrode is positively charged (anodal) and one is negatively charged (cathodal). Current is suggested to flow from the anodal to the cathodal electrode, through superficial cortical areas in the brain, inducing changes in membrane polarisation (Rozisky, Antunes, Brietzke, de Sousa, & Caumo, 2016; see Figure 1 for visual depiction). Anodal stimulation is widely believed to cause depolarisation of membrane potentials, resulting in a higher probability of an action potential (excitatory stimulation) and cathodal stimulation should hyperpolarise membrane potentials, reducing the likelihood of an action potential (Ardolino, Bossi, Barbieri, & Priori, 2005; Boros, Poreisz, Munchau, Paulus, & Nitsche, 2008; Nitsche & Paulus, 2000). Thus, it is important to think about the precise location of placement of the electrodes on the scalp in order to optimise the desired effects on brain polarisation.

Changes in neuronal excitability and behaviour have been shown to persist after stimulation and are often demonstrated by investigating the size of motor-evoked potentials (MEPs) elicited by TMS (see Figure 2 for demonstration of anodal and cathodal effects on motor cortex excitability). This raised the question as to whether synaptic plasticity is involved in the after-effects of stimulation. Although the precise mechanism of these changes is still not fully understood, stimulation is believed to influence protein synthesis and to cause

changes in intracellular calcium (Ca^{2+}) and sodium (Na^{+}), leading to changes in neurotransmission (Islam, Aftabuddin, Moriwaki, Hattori, & Hori, 1995). These are important aspects of long-term potentiation (LTP) and long-term depression (LTD; Monte-Silva et al., 2013) and thus, synaptic changes following tDCS have been likened to those involved in LTP and LTD. LTP describes the way in which connections between neurons are strengthened, following the Hebbian idea that 'cells that fire together wire together', creating long-lasting synaptic connections. LTD is opposite to LTP and leads to the weakening of synaptic connections as a result of neurons firing out of sync. LTP and LTD are considered to be the key processes involved in learning and memory.

LTP is dependent on N-methyl-D-aspartate (NMDA) receptors and direct current stimulation applied to the rat motor cortex has been found to induce (Fritsch et al., 2010; Podda et al., 2016), as well as modulate (Ranieri et al., 2012), NMDA receptor-dependent LTP. The role of NMDA activity has been further confirmed using pharmacological intervention. The excitatory effects of anodal tDCS were prolonged following a boost in NMDA receptor activity using the partial agonist D-cycloserine (Nitsche et al., 2004), and the effects of anodal and cathodal tDCS-induced plasticity were suppressed following administration of the NMDA receptor blocker dextromethorphan (Liebetanz, Nitsche, Tergau, & Paulus, 2002; Nitsche et al., 2003a). Thus, due to the clearly important role for NMDA in this process, it can be assumed that tDCS induces plasticity in the glutamatergic system. This system is also dependent on concentrations of calcium and sodium ions in the postsynaptic membrane and there is a body of in vitro animal evidence of increased levels of these ions following anodal direct current stimulation (Islam et al., 1995), as well as suppression of the effects of anodal direct current stimulation following administration of calcium and sodium channel blockers (Liebetanz et al., 2002; Nitsche et al., 2003a). This

provides evidence for the effects of stimulation on resting membrane potentials, with increases in these ion concentrations leading to enhanced synaptic transmission.

In combination with this, neurotransmission is also thought to be influenced via the GABAergic system, whereby anodal stimulation has been shown to reduce concentrations of GABA, an inhibitory neurotransmitter, therefore facilitating neuroplasticity (Nitsche et al., 2004; Stagg et al., 2009). Thus, it appears anodal stimulation acts on both glutamatergic and GABAergic neurotransmission in the brain, whereas cathodal simulation may act on the glutamatergic system. There are, however, reports of tDCS modulating dopaminergic and serotonergic activity and also varied effects on each of these systems depending on the precise regions, intensity of stimulation and neuronal cells targeted by the stimulation (Medeiros et al., 2012).

Finally, there is now evidence that tDCS may alter oscillatory activity in the underlying cortex, with enhanced beta and gamma oscillatory activity found in the visual cortex following anodal stimulation and diminished activity in these wavebands following cathodal stimulation (Antal, Kincses, Nitsche, Bartfai, & Paulus, 2004; Ardolino et al., 2005).

In conclusion, much work is still needed to fully elucidate the precise physiological mechanisms of action of anodal and cathodal tDCS on the brain and how these may differ as key stimulation parameters are altered. However, the aforementioned research does suggest that tDCS brings about changes in cortical excitability in the underlying cortex as well as influencing some aspects of neural plasticity, via a process similar to that of LTP and LTD.

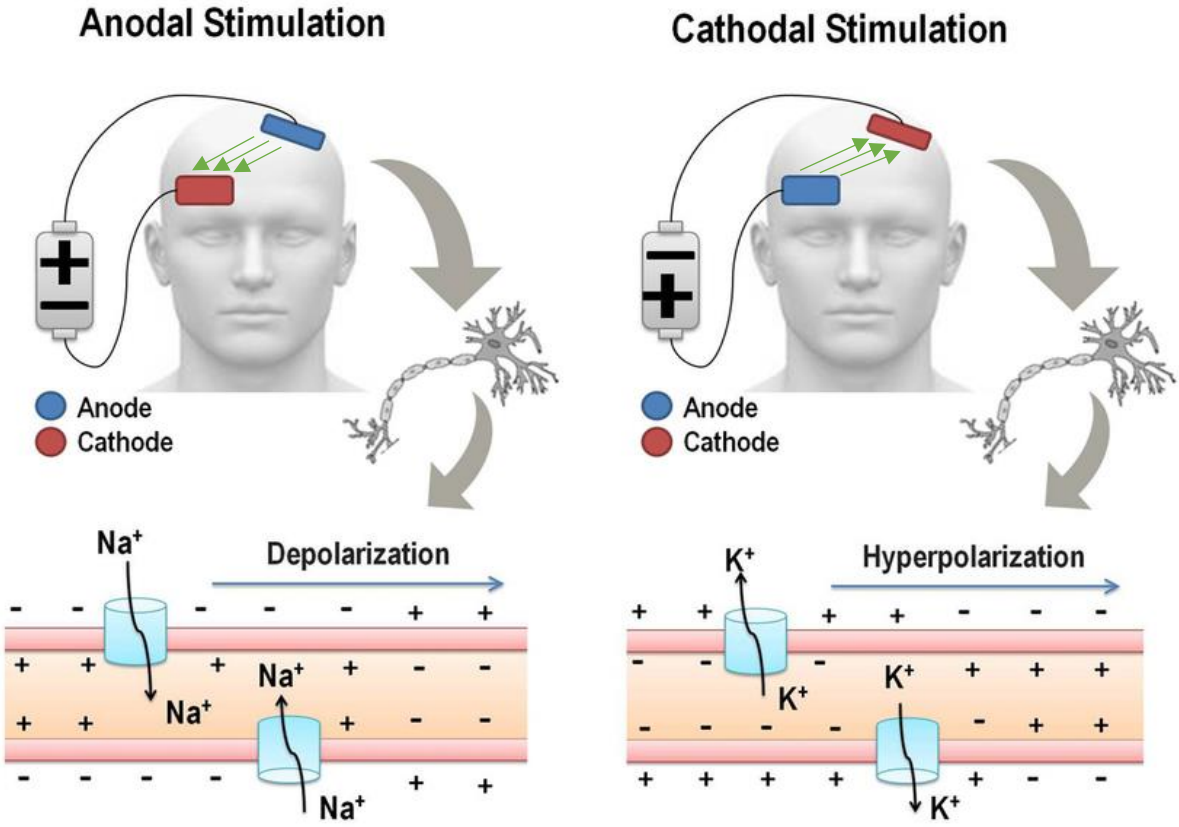

Figure 1. Illustration of tDCS setup and the effects of stimulation on membrane polarisation following anodal and cathodal stimulation. Green arrows represent direction of current flow from the anode to the cathode. Adapted from Rozisky et al. (2016).

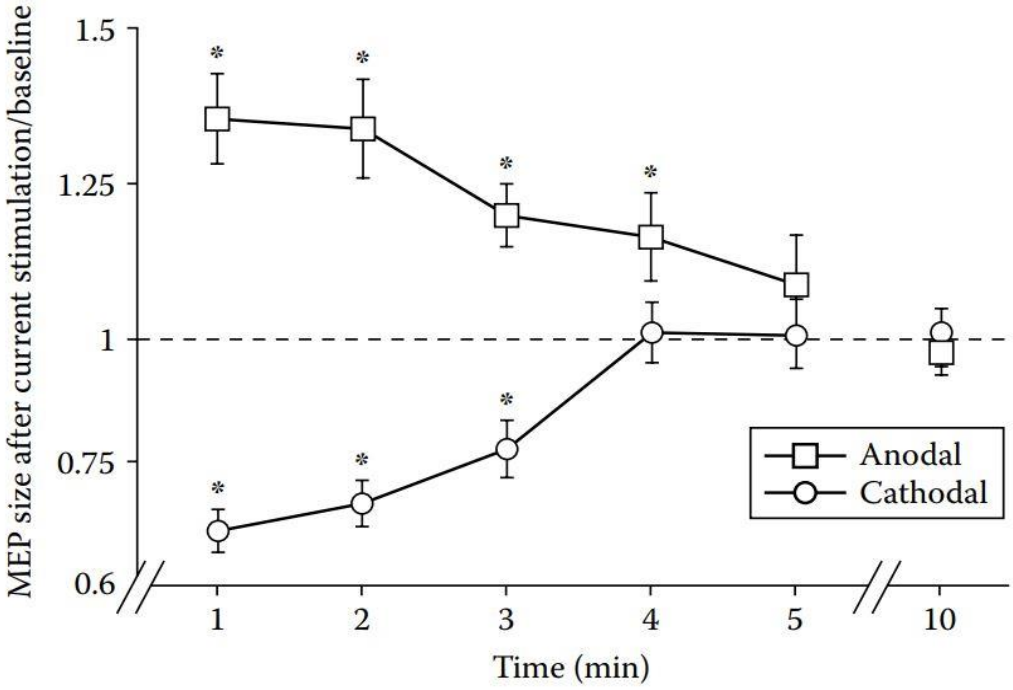

Figure 2. The time course of polarity-specific changes in motor cortex excitability for 5 minutes immediately following 5 minutes of tDCS at 1 mA to the motor cortex. Asterisks indicate significant differences in MEP amplitudes between anodal and cathodal conditions. Taken from Nitsche and Paulus (2001).

4.3. Methodological Issues

4.3.1. Stimulation parameters. There are a number of factors which affect the efficacy of tDCS, including the current intensity and density, length of stimulation, electrode placement, and whether the stimulation is delivered online or offline. These are important factors to consider to optimise the efficacy of tDCS and when interpreting results from studies using tDCS.

4.3.1.1. Stimulation intensity and electrode size. Current density is a proportionate measure of the stimulation intensity to the size of the electrode used to deliver the stimulation and can have a bearing on the physiological and behavioural effects of tDCS. The larger the surface area of the electrode, the more distributed the current is across the area underneath the electrode. Thus, electrode size should be taken into account when deciding the intensity of stimulation required to the relevant brain region of interest. Typically, tDCS studies utilise electrodes of 5 cm X 7 cm (35 cm^2) and deliver a current of between 1 and 2 mA, resulting in current densities between 0.029 and 0.086 mA/cm^2 (Paulus, Antal, & Nitsche, 2012). Although the peak current density is found in the cortex directly underlying the electrodes, with a rapid decrease in areas immediately neighbouring these (Lang et al., 2005), the use of such large electrodes does contribute to the low spatial resolution inherent with the use of tDCS, as the focality of the stimulation is compromised. The benefit, however, of using the 35 cm^2 electrode as opposed to much smaller electrodes is that it reduces the risk of skin burns caused by electricity concentrations on the scalp underneath the electrodes (Furubayashi et al., 2008).

When using 35 cm^2 electrodes, increases in current intensity from 0.2 mA to 1 mA have been investigated to gauge the lowest current intensity to elicit after-effects. When delivered over the motor cortex, a stimulation intensity of 0.4 mA was sufficient to see

observable effects on MEP amplitudes (Nitsche & Paulus, 2000). Moreover, current intensities of 3 mA or above have been found to cause some discomfort (Furubayashi et al., 2008). Thus, it is common safe practise across tDCS studies to utilise intensities of between 1 and 2 mA delivered by electrodes sized between 20 and 35 cm^2.

4.3.1.2. Stimulation duration. Effects of stimulation on behaviour are relatively short-lived, but after-effects have been found to increase with increasing durations of stimulation (Nitsche & Paulus, 2000). Nitsche and Paulus (2000) investigated motor cortex excitability following tDCS, as indexed by the amplitude of TMS-induced MEPs. The shortest stimulation duration for which measurable after-effects were found was 3 minutes at 1 mA. However, they showed a steep, non-linear increase in the duration of after-effects with increasing stimulation duration, with anodal tDCS for 5 minutes leading to after-effects for less than 5 minutes post stimulation, and stimulation for 13 minutes resulting in an over-proportionate increase of cortical excitability for up to 120 minutes (Figure 3). There is also some evidence to suggest repeated sessions of anodal tDCS may lead to subtle behavioural changes for up to several weeks (Nitsche & Paulus, 2000, 2001; Nitsche et al., 2003b). However, caution should be exercised here as there are some reports that increasing stimulation duration from 13 to 26 minutes converts the initial excitatory effect to an inhibitory effect (Monte-Silva et al., 2013). Thus, it seems there is an optimum stimulation duration for one session tDCS to optimise behavioural after-effects. In order to prolong the excitatory effects of anodal stimulation past a 90-120 minute window, it has been proposed that an interval of roughly 24 hours is required between repeated stimulation sessions (Reis et al., 2009). Similar increased after-effects with increased stimulation duration have been observed for cathodal stimulation; as stimulation duration increased from 9 to 18 minutes, so did the duration of the inhibitory after-effects observed on motor cortex-elicited MEPs (Monte-Silva et al., 2009).

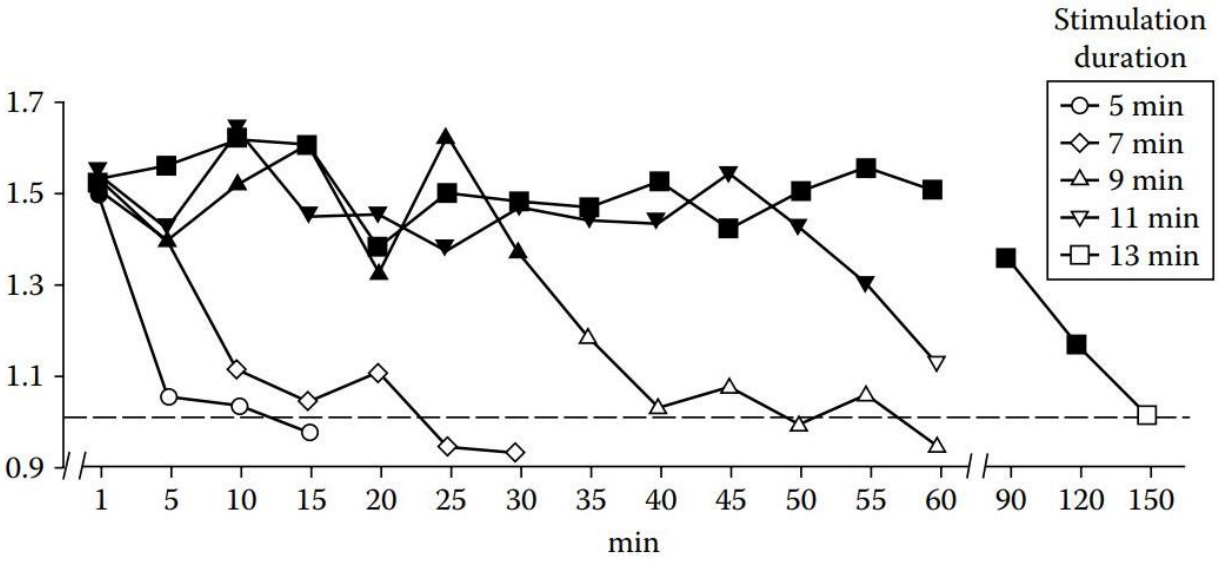

Figure 3. The time course of excitatory effects on the motor cortex (as indexed by TMS-induced MEP amplitudes) following different durations of anodal tDCS to the motor cortex. With increasing stimulation duration, stimulation effects are prolonged, lasting up to 120 minutes following 13 minutes of stimulation. Baseline excitability is represented by the dotted line and filled shapes represent significant changes compared to baseline. Taken from Nitsche and Paulus (2001).

4.3.1.3. Electrode positioning. The positioning of the anodal and cathodal electrodes during tDCS is particularly important when considering the physiological effects on the brain. As we know the current passes through cortical areas from the anodal to the cathodal electrode, it is important to appreciate that the modulation of excitability will be distributed over a wider area of the brain than simply the cortical area directly below the electrodes. Thus, it may be important to consider how a network of brain regions may be involved in a particular cognitive function of interest and this should be taken into account when setting stimulation parameters. One way to reduce the current density delivered to the cortical area under the reference electrode, and to maximise stimulation power over the area of interest relative to the area under the reference, is to significantly increase the size of the reference electrode (Nitsche et al., 2007).

There are a variety of electrode montages which can be used in a tDCS setup. The electrode of interest is placed over the cortical area of interest. While in some montages the reference electrode may be placed on a cephalic region (usually over a cortical area contralateral, or posterior, to the electrode of interest), in others it is placed on an extracephalic region such as the shoulder or upper arm, or on the forehead (Datta, Baker, Bikson, & Fridriksson, 2011). To demonstrate the influence of different electrode montages on tDCS-induced cortical excitability; during motor cortex stimulation, only two of six different electrode position combinations tested were shown to modulate motor cortex excitability (Nitsche et al., 2008), which is suggested to be because the different position combinations may modulate excitability of different neuronal populations (Priori et al., 1998). Moreover, during anodal stimulation to the motor cortex, placing the reference (cathodal) electrode over a cortical area contralateral to the anodal electrode resulted in excitatory after-effects on MEP amplitudes (Nitsche & Paulus, 2000), whereas when placed on the chin, inhibitory after-effects were observed (Priori et al., 1998).

Additionally, similar differences have been observed when the primary visual cortex was stimulated, whereby the reference electrode placement over the vertex or on the neck resulted in different after-effects on visual-evoked potentials in two different studies (Accornero, Voti, La Riccia, & Gregori, 2007; Antal et al., 2004). Other than the placement of the reference electrode, the methodology of these two sets of studies were equivalent, highlighting the vital importance of electrode positioning in determining the effects of stimulation on the brain.

Finally, Moliadze, Antal, and Paulus (2010) also demonstrated the importance of the distance between the two electrodes. They demonstrated that stimulation after-effects reduced significantly with increasing distance of the reference electrode from the anodal electrode during anodal stimulation to the motor cortex. In order to produce comparable after-effects

across each distance condition, stimulation intensity had to be adapted, with higher stimulation intensities required with greater distance between the anodal and reference electrodes.

Thus, from this evidence we see that electrode positioning is crucial in determining many aspects of current flow during direct current stimulation, and therefore has a huge bearing on the effect of the stimulation on cortical excitability (Utz, Dimova, Oppenländer, & Kerkhoff, 2010). The experiment detailed in Chapter 5 of this thesis therefore utilises a tDCS electrode montage whereby the anodal electrode is placed over the area of interest and the reference (cathodal) electrode is placed over the vertex, with this montage being well established for the cognitive function of interest, with consistent effects on cortical excitability across a number of studies (Hogeveen et al., 2015; Santiesteban, Banissy, Catmur, & Bird, 2012, 2015).

4.3.1.4. Online vs. offline stimulation. A further parameter to consider with the tDCS setup is the time at which stimulation is delivered. Online stimulation is delivered simultaneous with the completion of an experimental task, whereas offline stimulation is delivered prior to task completion. Studies looking to enhance learning often deliver their stimulation online during a learning or training experiment. For example, Nitsche et al. (2003b) demonstrated how applying anodal stimulation online improved learning rates during a motor learning task, whereas 10 minutes of offline stimulation prior to the completion of the same learning task had no effect on learning rates (Kuo, Paulus, & Nische, 2006).

Stagg and colleagues (Stagg et al., 2009) compared 10 studies investigating stimulation over the motor cortex, 8 of which utilised online stimulation, and 2 utilising offline stimulation and found stronger modulation of motor cortex excitability in the online studies. However, when considering cognitive domains such as attention, a direct comparison

of performance on the same task was made between online and offline stimulation to the visual cortex, revealing greater facilitation in the offline condition. Commensurately, Hsu, Ku, Zanto, and Gazzaley (2015) presented a meta-analysis of the effects of tDCS on cognitive function in healthy individuals, finding offline stimulation to produce stronger enhancement of cognitive functioning than online stimulation. Interestingly, however, they found the reverse was true in individuals with Alzheimer's disease, whereby more pronounced improvements of cognitive functioning were observed following online compared to offline stimulation.

It appears that both online and offline designs may be successful in inducing after-effects. However, the efficacy of online and offline stimulation may be dependent on the study design, including the domain and even the target population of interest, with some behavioural functions benefitting from one more than the other. More controlled comparisons of online and offline designs are required, particularly within and between various cognitive domains in order to elucidate the optimal time to administer tDCS to elicit desired stimulation effects.

4.3.1.5. Sham vs. active control stimulation. Another important feature of any tDCS design is the use of a baseline or control stimulation condition to establish the specificity of the effects of stimulation, i.e. to identify whether the effects are specific to active stimulation (relative to a placebo effect) or specific to stimulation of a particular brain region. One method is to use active stimulation vs. sham (non-active) stimulation. For instance, the electrodes are placed at the same locations on the scalp, however, in the sham condition, the constant current stimulator is turned on for an initial 15-30 seconds, after which the stimulation ceases. This gives the initial tingling sensation experienced with active stimulation, making the physical experience as similar as possible between the conditions, and it is widely accepted that participants cannot distinguish sham and active stimulation

(Russo, Wallace, Fitzgerald, & Cooper, 2013). Despite this, there are reports of differences in the experience of active and sham stimulation (Kessler, Turkeltaub, Benson, & Hamilton, 2012). Moreover, a behavioural difference observed between the after-effects of sham and active stimulation does not rule out the possibility that active stimulation to any other region of the brain might also induce the same after-effects. Thus, sham stimulation by itself does not make for the best control condition because of its lack of specificity in assessing the involvement of a particular brain region.

Alternatively, a more experimentally rigorous control condition would involve an active control site whereby, for example, anodal stimulation is administered to a region of interest in one condition, and an area not thought to be involved in the behaviour of interest in another condition, whilst the reference electrode is placed over the same region in both conditions. Differences in after-effects found here allow conclusions to be made about the selective involvement of the area of interest, thus acting as a more informative control condition (Parkin, Ekhtiari, & Walsh, 2015).

4.3.2. Safety considerations. With the current surge in the use of tDCS in both research and clinical settings, it is important to consider the health and safety of the participants when designing stimulation protocols and carrying out stimulation. TDCS is widely accepted to be a safe, non-invasive form of brain stimulation, with the most severe adverse effects being limited to a few reports of small lesions on the skin underlying the electrodes (Frank et al., 2010; Palm et al., 2008; Rodriguez, Opisso, Pascual-Leone, & Soler, 2014). These were, however, reported in studies stimulating in the higher range of standard stimulation intensity (≥2 mA) and often from multiple sessions of stimulation.

Other minor adverse side effects reported in both healthy and clinical populations include slight skin irritation or phosphenes (bright flashes of light) at the start of stimulation,

as well as headache, nausea, dizziness, itching or tingling sensations, and reduced concentration after stimulation (Brunoni et al., 2011; Nitsche et al., 2003c; Poreisz, Boros, Antal, & Paulus, 2007). Poreisz et al. (2007) examined side effects across a number of studies conducted in healthy and clinical populations, with itching sensations being the most commonly reported side effect in healthy participants and headaches most commonly reported in patients. However, in a meta-analysis of clinical trials by Brunoni and colleagues (Brunoni et al., 2011), the most commonly reported side effects of itching or tingling were reported numerically, but not statistically significantly more frequently following active than non-active stimulation. Brunoni and colleagues do nonetheless describe a lack of consistency across studies in the tendency and means of reporting adverse side effects in tDCS studies, thus calling for stricter guidelines to improve our understanding of the safe parameters for tDCS in future research and clinical settings.

There have been also reports of current-induced tissue damage in the brains of rats following direct current stimulation (Liebetanz et al., 2009). However, as Nitsche et al. (2003c) identify, unlike some animal studies where stimulation may be applied directly to the brain tissue, human direct current stimulation is administered transcranially, with electrodes placed on the scalp and not in direct contact with the brain. Moreover, the current density used in this study was roughly two orders of magnitude higher than that typically used in human tDCS studies. Thus, it is unlikely that such adverse effects will be observed in humans (Nitsche & Paulus, 2011).

Current commonly used parameters are widely tested and approved to be safe with minimal side effects. Accordingly, stimulation parameters are set in line with these safety criteria and are usually kept to a very weak current between 0.5 and 2 mA (Bikson, Datta, & Elwassif, 2009). Moreover, various safety measures are taken to ensure that the parameters are safe for each individual undergoing stimulation. This includes participants completing a

thorough screening questionnaire. As tDCS involves electricity, it is not advisable for people who have a pacemaker, an implanted medication pump, a metal plate in the skull or metal objects inside the eye or skull (e.g. after brain surgery or a shrapnel wound). The other main concern most commonly associated with other forms of brain stimulation, such as TMS, is that in rare circumstances this method has induced seizures. However, tDCS modulates membrane potentials rather than neuronal excitability directly and therefore there have not been reports of this occurring with the use of tDCS parameters within the current accepted safe limits (Nitsche et al., 2003c). Nevertheless, individuals susceptible to seizures (i.e. with a history of epilepsy) and those taking neuroleptic medication should be prevented from taking part in tDCS studies.

4.3.3. Inconsistency in cathodal effects. It was the investigation of the effects of tDCS on motor function which lead to our understanding of the effects of anodal and cathodal stimulation on the brain. However, with the later addition of cognitive studies, these polarity effects have been shown not to replicate across some studies (Jacobson, Koslowsky, & Lavidor, 2012). These inconsistencies, however, have generally concerned cathodal stimulation showing effects on motor function but little or no observable effects on cognitive function, whereas behavioural effects of anodal stimulation appear to be consistently excitatory across both motor and cognitive domains. It is worth noting, however, that although the effects of cathodal stimulation may not be observable on behaviour, this does not rule out the possibility that it is having an inhibitory effect at the physiological and molecular level. Moreover, cathodal effects may be more complicated than those of anodal stimulation. For example, decreasing neuronal excitability in some regions of the brain during certain cognitive tasks may reduce neuronal competition and thus improve performance in some domains.

Neuronal excitability in areas performing various cognitive functioning is also not expected to be at zero prior to tDCS, and thus stimulation is being applied to an already active cortex (Silvanto, Muggleton, & Walsh, 2008). Therefore, cathodal effects on neuronal excitability may be more difficult to find, or may be smaller in magnitude, than those due to anodal stimulation. Anodal stimulation simply increases the already high neuronal excitability, whereas cathodal stimulation may need to be significantly greater in order to produce comparable inhibitory effects on excitability (Jacobson et al., 2012). In comparison, however, the motor cortex is thought to be less active at rest which might explain the more equivalent magnitudes of effects of anodal and cathodal stimulation on motor cortex activity but not on cognitive function.

Thus, a straightforward comparison of anodal and cathodal effects, as well as inferring neuronal excitability from cognitive performance may not be possible. It is important to establish how anodal and cathodal stimulation affects domains of behaviour differentially, and using well-researched stimulation parameters will allow for more concrete conclusions to be drawn from the results of tDCS studies. Consequently, in Chapter 5 of this thesis, the stimulation parameters have been set in line with those from previous studies investigating the role of the temporoparietal junction (TPJ) in self-other control processes (Hogeveen et al., 2015; Santiesteban et al., 2012, 2015). Crucially these studies showed excitatory effects on the behaviour of interest when the anodal electrode was placed over the TPJ, but not when placed over a mid-occipital control region (with the reference, cathodal electrode over the vertex), but no effects on behaviour when either sham stimulation was administered or when the cathodal electrode was placed over the TPJ (with the reference, anodal electrode placed over the vertex). Thus, the tDCS parameters described in Chapter 5 based on this previous research, allow a prediction as to the direction and specificity of the behavioural effects expected if self-other control recruits the TPJ.

4.3.4. Inconsistency in the neurophysiological and behavioural effects of tDCS.

Recently, the reliability and replicability of experimental findings within psychology have been called into question and this is also a growing concern in the area of neurostimulation, with the specific suggestion that tDCS effects are inconsistent and unreliable. It is now common to take a more meta-analytical approach to the study of the effects of tDCS on brain and cognition with the aim of revealing whether findings are consistent and reliable across stimulation parameters, cognitive domains and participant populations. For example, meta-analyses have been carried out across the whole spectrum of applications for tDCS, such as in modulating specific cognitive functions (Brunoni & Vanderhasselt, 2014; Hill, Fitzgerald, & Hoy, 2016; Kang, Summers, & Cauraugh, 2016; Khalighinejad, Di Costa, & Haggard, 2016; Price et al., 2015); in treating symptoms of specific clinical conditions (Berlim, Van den Eynde, & Daskalakis, 2013; Brunoni et al., 2016; Butler et al., 2013; Meron, Hedger, Garner, & Baldwin, 2015; Shiozawa et al., 2014); as well as meta-analyses of stimulation to specific brain regions (Dedoncker, Brunoni, Baeken, & Vanderhasselt, 2016; Donaldson, Rinehart, & Enticott, 2015); reviews to categorise the effects of different electrode montages (Nasseri, Nitsche, & Ekhtari, 2015), and finally reviews to try to find common effects of all variations of tDCS parameters across all behavioural outcomes (Horvath, Forte, & Carter, 2015a; Summers, Kang, & Cauraugh, 2016) and neurophysiological effects (Horvath, Forte, & Carter, 2015b; Medeiros et al., 2012).

Despite general consistency between findings in a number of these meta-analyses, a handful revealed mixed or no consistent results for the effect of tDCS on the brain and behaviour (Horvath et al., 2015a, 2015b). However, it has been suggested that a reason for such a mixed picture of the effects of tDCS, and the problem of comparing all studies directly, is the issue of the heterogeneity of the experimental protocols (Donaldson et al., 2015; Shiozawa et al., 2014). For instance, the more the meta-analysis collapses across

cognitive domains, brain regions, electrode montages and various other stimulation parameters, the more mixed the results are likely to be (Price & Hamilton, 2015). Those studies taking a more specific investigation of one facet of stimulation, such as a specific brain region, cognitive function or stimulation parameter (e.g. electrode montage, stimulation intensity or timing of stimulation), indeed find more consistent stimulation results (Brunoni & Vanderhasselt, 2014; Donaldson et al., 2015; Kang et al., 2016; Price et al., 2015; Sellaro et al., 2017).

Particularly, Horvath and colleagues argue for no reliable neurophysiological (Horvath et al, 2015b) or cognitive (Horvath et al., 2015a) effects of tDCS. However, it has since been argued that the methods of these meta-analyses were themselves questionable, with problems in their data selection and statistical approach used, as well as the methods of many studies being mischaracterised for the sake of grouping studies into arbitrary cognitive domains (Price & Hamilton, 2015). Additionally, experts in the field (Nitsche, Bikson, & Bestmann, 2015) have commented that meta-analyses aiming to find a binary conclusion of what direct current stimulation does to the brain and behaviour, or whether or not it 'works', are miscalibrated to the current state of the field. We know that a number of factors affect the expected effects of direct current stimulation on brain excitability and behaviour, including current intensity, electrode size, the cognitive function being investigated and probably most importantly, the electrode positioning and montage of the two electrodes. Thus, although large scale meta-analyses are preferable and improve power to detect effects, a more realistic place to begin is with straight forward replications as well as more systematic choice of parameters and data analyses where factors influencing stimulation effects are identified and controlled. This will allow for future establishment of more reliable, consistent and targeted stimulation from which our understanding of the physiological and behavioural mechanisms will be improved.

Finally, an important consideration in determining the reliability of the neurophysiological and behavioural effects of brain stimulation is the presence of publication bias. This can be explained as the outcome of an experiment having an impact on whether the research is published in an academic journal. Publication bias is generally seen via a positive bias to publish only results which are significant and may be introduced by the experimenter failing to publish negative findings or academic journals accepting more manuscripts presenting positive than negative findings (Dickersin, 1990). This presents a problem for the progression of science, as negative findings are as important in advancing knowledge and understanding as positive findings, but are significantly underrepresented in academic journals (Jooper, Schmitz, Annable, & Boksa, 2012). The problem of positive publication bias is reported to be a particular issue in the field of neurostimulation. For example, Héroux, Loo, Taylor, & Gandevia (2017) recently demonstrated that 154 researchers (from an invited 976) utilising brain stimulation in their research have a median of 3 published brain stimulation studies and 2 unpublished, whilst only 45-50% reported being able to routinely reproduce published results.

However, the introduction of pre-registration of scientific protocols and subsequent publication regardless of the direction of the findings, as well as the development of methods to account for publication bias in meta-analyses of existing published research (Duval & Tweedie, 2000; Peters, Sutton, Jones, Abrams, & Rushton, 2006) can help to alleviate these concerns in the future. This is particularly important in the field of neurostimulation, in which publication of null findings is fundamental to advancing our understanding of the precise effects of each form of stimulation on the brain and cognition.

4.3.5. Individual differences. Finally, an important issue to consider in evaluating the consistency of effects of tDCS is individual differences both between and within individuals in their response to tDCS (Chew, Ho, & Loo, 2015; Fertonani & Miniussi, 2017;

Krause & Cohen Kadosh, 2014). For example, individual differences in cortical excitability may contribute to differing effects of stimulation, and these can vary according to factors such as age and gender between participants but also pre-existing brain states and hormone levels within individuals (Falkenberg, Westerhausen, Specht, & Hugdahl, 2012; Krause & Cohen Kadosh, 2014). Some factors have even been shown to reverse the polarity of stimulation effects (Hsu, Juan, & Tseng, 2016; Krause, Márquez-Ruiz, & Cohen Kadosh 2013). For example, Krause and colleagues (Krause et al., 2013) demonstrated that vastly different effects of anodal tDCS may be found between individuals with high or low regional excitability or even between brain regions within the same individual with high or low regional excitability. The effects of stimulation on any one individual have however, been reported to be stable across testing sessions (López-Alonso, Fernández-del-Olmo, Costantini, Gonzalez-Henriquez, & Cheeran, 2015).

It has been suggested that across a spectrum of levels of regional excitability, stimulation effects (excitatory or inhibitory) may follow an inverted U-shaped pattern, whereby there is an optimum level of excitability to maximise the effect of stimulation. Thus, it may be possible, and perhaps more informative to our understanding, to split a sample into 'responders' and 'non-responders' and to investigate the factors which influence this (López-Alonso, Cheeran, Río-Rodríguez, & Fernández-Del-Olmo, 2014). It may also be crucial in future studies to take a more individual differences approach to effects of stimulation across an experiment, rather than examining grand averages of physiological and behavioural effects of stimulation for each person. This is demonstrated by Hamada, Murase, Hasan, Balaratnam, and Rothwell (2013), who tracked the patterns of MEP amplitudes in their participants across 8 time points during an experiment. Participants were highly variable but showed definite physiological effects of the stimulation, which were lost when amplitudes were averaged across time points.

4.4. Conclusions

TDCS has an extensive and varied history in its development and uses in research and clinical practice, providing some of the best understanding we have to date regarding brain physiology investigated in a causal manner, but without long-lasting damage to the brain. Despite the recent, potentially miscalibrated, discussion regarding a binary conclusion for what tDCS does to the brain and behaviour, it seems clear that tDCS has huge potential to reveal the complexities of the functioning of the human brain. For example, the number of stimulation parameters which can be manipulated with tDCS, including electrode placement montages, timing, frequency and intensity of stimulation and the behavioural outcome of interest allow researchers to design very specific stimulation protocols to probe at highly-specialised questions about brain function. There is still much work to be done to fully elucidate the precise effects of varying each stimulation parameter at a physiological and behavioural level. These include the use of more evidence-based, rigorously controlled protocols, replication studies to establish reliable and consistent protocols, as well as well-categorised meta-analytic investigations of each cognitive function and different combinations of stimulation parameters.

4.5 References

Accornero, N., Li Voti, P., La Riccia, M., & Gregori, B. (2007). Visual evoked potentials modulation during direct current cortical polarization. *Experimental Brain Research, 178,* 261-266.

Albert, D. J. (1966). The effects of polarizing currents on the consolidation of learning. *Neuropsychologia, 4,* 65-77.

Antal, A., Kincses, T. Z., Nitsche, M. A., Bartfai, O., & Paulus, W. (2004). Excitability changes induced in the human primary visual cortex by transcranial direct current stimulation: Direct electrophysiological evidence. *Investigative Ophthalmology & Visual Science, 45,* 702-707.

Ardolino, G., Bossi, B., Barbieri, S., & Priori, A. (2005). Non- synaptic mechanisms underlie the after- effects of cathodal transcutaneous direct current stimulation of the human brain. *The Journal of Physiology, 568,* 653-663.

Berlim, M. T., Van den Eynde, F., & Daskalakis, Z. J. (2013). Clinical utility of transcranial direct current stimulation (tDCS) for treating major depression: A systematic review and meta-analysis of randomized, double-blind and sham-controlled trials. *Journal of Psychiatric Research, 47,* 1-7.

Bikson, M., Datta, A., & Elwassif, M. (2009). Establishing safety limits for transcranial direct current stimulation. *Clinical Neurophysiology, 120,* 1033-1034.

Borckardt, J. J., Bikson, M., Frohman, H., Reeves, S. T., Datta, A., Bansal, V., ... & George, M. S. (2012). A pilot study of the tolerability and effects of high-definition transcranial direct current stimulation (HD-tDCS) on pain perception. *The Journal of Pain, 13,* 112-120.

Borckardt, J. J., Romagnuolo, J., Reeves, S. T., Madan, A., Frohman, H., Beam, W., & George, M. S. (2011). Feasibility, safety, and effectiveness of transcranial direct current stimulation for decreasing post-ERCP pain: A randomized, sham-controlled, pilot study. *Gastrointestinal Endoscopy*, *73*, 1158-1164.

Boros, K., Poreisz, C., Munchau, A., Paulus, W., & Nitsche, M. A. (2008). Premotor transcranial direct current stimulation (tDCS) affects primary motor excitability in humans. *The European Journal of Neuroscience*, *27*, 1292–1300.

Brunoni, A. R., Amadera, J., Berbel, B., Volz, M. S., Rizzerio, B. G., & Fregni, F. (2011). A systematic review on reporting and assessment of adverse effects associated with transcranial direct current stimulation. *International Journal of Neuropsychopharmacology*, *14*, 1133-1145.

Brunoni, A. R., Moffa, A. H., Fregni, F., Palm, U., Padberg, F., Blumberger, D. M., ... & Loo, C. K. (2016). Transcranial direct current stimulation for acute major depressive episodes: meta-analysis of individual patient data. *The British Journal of Psychiatry*, *208*, 522-531.

Brunoni, A. R., & Vanderhasselt, M. A. (2014). Working memory improvement with non-invasive brain stimulation of the dorsolateral prefrontal cortex: A systematic review and meta-analysis. *Brain and Cognition*, *86*, 1-9.

Butler, A. J., Shuster, M., O'hara, E., Hurley, K., Middlebrooks, D., & Guilkey, K. (2013). A meta-analysis of the efficacy of anodal transcranial direct current stimulation for upper limb motor recovery in stroke survivors. *Journal of Hand Therapy*, *26*, 162-171.

Chew, T., Ho, K. A., & Loo, C. K. (2015). Inter-and intra-individual variability in response to transcranial direct current stimulation (tDCS) at varying current intensities. *Brain Stimulation*, *8*, 1130-1137.

Creutzfeldt, O. D., Fromm, G. H., & Kapp, H. (1962). Influence of transcortical dc currents on cortical neuronal activity. *Experimental Neurology*, *5*, 436-452.

DaSilva, A. F., Volz, M. S., Bikson, M., & Fregni, F. (2011). Electrode Positioning and Montage in Transcranial Direct Current Stimulation. *Journal of Visualized Experiments*, e2744.

Datta, A., Baker, J. M., Bikson, M., & Fridriksson, J. (2011). Individualized model predicts brain current flow during transcranial direct-current stimulation treatment in responsive stroke patient. *Brain Stimulation*, *4*, 169-174.

Debru, A. (2006). The power of torpedo fish as a pathological model to the understanding of nervous transmission in antiquity. *Comptes Rendus Biologies*, *329*, 298-302.

Demirtas-Tatlidede, A., Vahabzadeh-Hagh, A. M., & Pascual-Leone, A. (2013). Can noninvasive brain stimulation enhance cognition in neuropsychiatric disorders?. *Neuropharmacology*, *64*, 566-578.

Dedoncker, J., Brunoni, A. R., Baeken, C., & Vanderhasselt, M. A. (2016). The effect of the interval-between-sessions on prefrontal transcranial direct current stimulation (tDCS) on cognitive outcomes: A systematic review and meta-analysis. *Journal of Neural Transmission*, *123*, 1159-1172.

Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *Jama*, *263*, 1385-1389.

Donaldson, P. H., Rinehart, N. J., & Enticott, P. G. (2015). Noninvasive stimulation of the temporoparietal junction: A systematic review. *Neuroscience & Biobehavioral Reviews*, *55*, 547-572.

Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel- plot–based method of testing and adjusting for publication bias in meta- analysis. *Biometrics*, *56*, 455-463.

Falkenberg, L. E., Westerhausen, R., Specht, K., & Hugdahl, K. (2012). Resting-state glutamate level in the anterior cingulate predicts blood-oxygen level-dependent response to cognitive control. *Proceedings of the National Academy of Sciences*, *109*, 5069-5073.

Fertonani, A., & Miniussi, C. (2017). Transcranial electrical stimulation: what we know and do not know about mechanisms. *The Neuroscientist*, *23*, 109-123.

Frank, E., Wilfurth, S., Landgrebe, M., Eichhammer, P., Hajak, G., & Langguth, B. (2010). Anodal skin lesions after treatment with transcranial direct current stimulation. *Brain Stimulation*, *3*, 58-59

Fregni, F., Boggio, P. S., Nitsche, M., Bermpohl, F., Antal, A., Feredoes, E., ... & Pascual-Leone, A. (2005). Anodal transcranial direct current stimulation of prefrontal cortex enhances working memory. *Experimental Brain Research*, *166*, 23-30.

Fregni, F., Boggio, P. S., Santos, M. C., Lima, M., Vieira, A. L., Rigonatti, S. P., … & Pascual-Leone, A. (2006). Noninvasive cortical stimulation with transcranial direct current stimulation in parkinson's disease. *Movement Disorders: Official Journal of the Movement Disorder Society*, *21*, 1693-702.

Fritsch, B., Reis, J., Martinowich, K., Schambra, H. M., Ji, Y., Cohen, L. G., & Lu, B. (2010). Direct current stimulation promotes BDNF-dependent synaptic plasticity: Potential implications for motor learning. *Neuron*, *66*, 198-204.

Furubayashi, T., Terao, Y., Arai, N., Okabe, S., Mochizuki, H., Hanajima, R., ... & Ugawa, Y. (2008). Short and long duration transcranial direct current stimulation (tDCS) over the human hand motor area. *Experimental Brain Research*, *185*, 279-286.

Gill, J., Shah-Basak, P. P., & Hamilton, R. (2015). It's the thought that counts: Examining the task-dependent effects of transcranial direct current stimulation on executive function. *Brain Stimulation*, *8*, 253-259.

Hamada, M., Murase, N., Hasan, A., Balaratnam, M., and Rothwell, J. C. (2013). The role of interneuron networks in driving human motor cortical plasticity. *Cerebral Cortex*, *23*, 1593–1605.

Héroux, M. E., Loo, C. K., Taylor, J. L., & Gandevia, S. C. (2017). Questionable science and reproducibility in electrical brain stimulation research. *PloS One*, *12*, e0175635.

Hill, A. T., Fitzgerald, P. B., & Hoy, K. E. (2016). Effects of anodal transcranial direct current stimulation on working memory: A systematic review and meta-analysis of findings from healthy and neuropsychiatric populations. *Brain Stimulation*, *9*, 197-208.

Hogeveen, J., Obhi, S. S., Banissy, M. J., Santiesteban, I., Press, C., Catmur, C., & Bird, G. (2015). Task-dependent and distinct roles of the temporoparietal junction and inferior frontal cortex in the control of imitation. *Social Cognitive and Affective Neuroscience*, *10*, 1003-1009.

Horvath, J. C., Forte, J. D., & Carter, O. (2015a). Quantitative review finds no evidence of cognitive effects in healthy populations from single-session transcranial direct current stimulation (tDCS). *Brain Stimulation*, *8*, 535-550.

Horvath, J. C., Forte, J. D., & Carter, O. (2015b). Evidence that transcranial direct current stimulation (tDCS) generates little-to-no reliable neurophysiologic effect beyond MEP amplitude modulation in healthy human subjects: A systematic review. *Neuropsychologia*, *66*, 213-236.

Hsu, T. Y., Juan, C. H., & Tseng, P. (2016). Individual differences and state-dependent responses in transcranial direct current stimulation. *Frontiers in Human Neuroscience*, *10*, 643.

Hsu, W. Y., Ku, Y., Zanto, T. P., & Gazzaley, A. (2015). Effects of noninvasive brain stimulation on cognitive function in healthy aging and Alzheimer's disease: A systematic review and meta-analysis. *Neurobiology of Aging*, *36*, 2348-2359.

Islam, N., Aftabuddin, M., Moriwaki, A., Hattori, Y., & Hori, Y. (1995). Increase in the calcium level following anodal polarization in the rat brain. *Brain Research*, *684*, 206-208.

Jacobson, L., Koslowsky, M., & Lavidor, M. (2012). TDCS polarity effects in motor and cognitive domains: A meta-analytical review. *Experimental Brain Research.*, *216*, 1-10.

Joober, R., Schmitz, N., Annable, L., & Boksa, P. (2012). Publication bias: what are the challenges and can they be overcome?. *Journal of Psychiatry & Neuroscience*, 37, 149-152.

Kang, N., Summers, J. J., & Cauraugh, J. H. (2016). Transcranial direct current stimulation facilitates motor learning post-stroke: A systematic review and meta-analysis. *Journal of Neurology, Neurosurgery, and Psychiatry*, *87*, 345-355.

Kellaway, P. (1946). The part played by electric fish in the early history of bioelectricity and electrotherapy. *Bulletin of the History of Medicine*, *20*, 112-137.

Kessler, S. K., Turkeltaub, P. E., Benson, J. G., & Hamilton, R. H. (2012). Differences in the experience of active and sham transcranial direct current stimulation. *Brain Stimulation*, *5*, 155-162.

Khalighinejad, N., Di Costa, S., & Haggard, P. (2016). Endogenous action selection processes in dorsolateral prefrontal cortex contribute to sense of agency: A meta-analysis of tDCS studies of 'intentional binding'. *Brain Stimulation*, *9*, 372-379.

Krause, B., & Cohen Kadosh, R. (2014). Not all brains are created equal: The relevance of individual differences in responsiveness to transcranial electrical stimulation. *Frontiers in Systems Neuroscience*, *8*, 25-25.

Krause, B., Márquez-Ruiz, J., & Cohen Kadosh, R. (2013). The effect of transcranial direct current stimulation: A role for cortical excitation/inhibition balance?. *Frontiers in Human Neuroscience*, *7*, 602.

Kuo, M. -F., Paulus, W., & Nitsche, M. A. (2006). Sex differences in cortical neuroplasticity in humans. *Neuroreport*, *17*, 1703-1707.

Lang, N., Siebner, H. R., Ward, N. S., Lee, L., Nitsche, M. A., Paulus, W., ... & Frackowiak, R. S. (2005). How does transcranial DC stimulation of the primary motor cortex alter regional neuronal activity in the human brain?. *European Journal of Neuroscience*, *22*, 495-504.

Liebetanz, D., Koch, R., Mayenfels, S., König, F., Paulus, W., & Nitsche, M. A. (2009). Safety limits of cathodal transcranial direct current stimulation in rats. *Clinical Neurophysiology*, *120*, 1161-1167.

Liebetanz, D., Nitsche, M. A., Tergau, F., & Paulus, W. (2002). Pharmacological approach to the mechanisms of transcranial DC- stimulation- induced after- effects of human motor cortex excitability. *Brain*, *125*, 2238-2247.

Lindenberg, R., Renga, V., Zhu, L. L., Nair, D., & Schlaug, G. M. D. P. (2010). Bihemispheric brain stimulation facilitates motor recovery in chronic stroke patients. *Neurology*, *75*, 2176-2184.

Lippold, O. C. J., & Redfearn, J. W. T. (1964). Mental changes resulting from the passage of small direct currents through the human brain. *The British Journal of Psychiatry*, *110*, 768-772.

Lolas, F. (1977). Brain polarization: Behavioral and therapeutic effects. *Biological Psychiatry*, *12*, 37-47.

López-Alonso, V., Cheeran, B., Río-Rodríguez, D., & Fernández-del-Olmo, M. (2014). Inter-individual variability in response to non-invasive brain stimulation paradigms. *Brain Stimulation*, *7*, 372-380.

López-Alonso, V., Fernández-del-Olmo, M., Costantini, A., Gonzalez-Henriquez, J. J., & Cheeran, B. (2015). Intra-individual variability in the response to anodal transcranial direct current stimulation. *Clinical Neurophysiology*, *126*, 2342-2347.

Medeiros, L. F., de Souza, I. C. C., Vidor, L. P., de Souza, A., Deitos, A., Volz, M. S., ... & Torres, I. L. (2012). Neurobiological effects of transcranial direct current stimulation: a review. *Frontiers in Psychiatry*, *3*, 110.

Meron, D., Hedger, N., Garner, M., & Baldwin, D. S. (2015). Transcranial direct current stimulation (tDCS) in the treatment of depression: Systematic review and meta-analysis of efficacy and tolerability. *Neuroscience & Biobehavioral Reviews, 57*, 46-62.

Moliadze, V., Antal, A., & Paulus, W. (2010). Electrode-distance dependent after-effects of transcranial direct and random noise stimulation with extracephalic reference electrodes. *Clinical Neurophysiology, 121*, 2165-2171.

Monte-Silva, K., Kuo, M. F., Hessenthaler, S., Fresnoza, S., Liebetanz, D., Paulus, W., & Nitsche, M. A. (2013). Induction of late LTP-like plasticity in the human motor cortex by repeated non-invasive brain stimulation. *Brain Stimulation, 6*, 424-432.

Monte-Silva, K., Kuo, M. F., Thirugnanasambandam, N., Liebetanz, D., Paulus, W., & Nitsche, M. A. (2009). Dose-dependent inverted U-shaped effect of dopamine (D2-like) receptor activation on focal and nonfocal plasticity in humans. *Journal of Neuroscience*, *29*, 6124-6131.

Monti, A., Ferrucci, R., Fumagalli, M., Mameli, F., Cogiamanian, F., Ardolino, G., & Priori, A. (2013). Transcranial direct current stimulation (tDCS) and language. *Journal of Neurology, Neurosurgery & Psychiatry*, *84*, 832-842.

Nasseri, P., Nitsche, M. A., & Ekhtiari, H. (2015). A framework for categorizing electrode montages in transcranial direct current stimulation. *Frontiers in Human Neuroscience, 9*, 54.

Nitsche, M. A., Bikson, M., & Bestmann, S. (2015). On the use of meta-analysis in neuromodulatory non-invasive brain stimulation. *Brain Stimulation, 3*, 666-667.

Nitsche, M. A., Boggio, P. S., Fregni, F., & Pascual-Leone, A. (2009). Treatment of depression with transcranial direct current stimulation (tDCS): A review. *Experimental Neurology*, *219*, 14-19.

Nitsche, M. A., Cohen, L. G., Wassermann, E. M., Priori, A., Lang, N., Antal, A., ... & Pascual-Leone, A. (2008). Transcranial direct current stimulation: State of the art 2008. *Brain Stimulation*, *1*, 206-223.

Nitsche, M. A., Doemkes, S., Karakoese, T., Antal, A., Liebetanz, D., Lang, N., ... & Paulus, W. (2007). Shaping the effects of transcranial direct current stimulation of the human motor cortex. *Journal of Neurophysiology*, *97*, 3109-3117.

Nitsche, M. A., Fricke, K., Henschke, U., Schlitterlau, A., Liebetanz, D., Lang, N., ... & Paulus, W. (2003a). Pharmacological modulation of cortical excitability shifts induced by transcranial direct current stimulation in humans. *The Journal of Physiology*, *553*, 293-301.

Nitsche, M. A., Jaussi, W., Liebetanz, D., Lang, N., Tergau, F., & Paulus, W. (2004). Consolidation of human motor cortical neuroplasticity by D-cycloserine. *Neuropsychoharmacology*, *29*, 1573-1578.

Nitsche, M. A., Liebetanz, D., Lang, I. N., Antal, A., Tergau, F., & Paulus, W. (2003c). "Safety criteria for transcranial direct current stimulation (tDCS) in humans". *Clinical Neurophysiology*, *114*, 2220–2222.

Nitsche, M. A., Liebetanz, D., Schlitterlau, A., Henschke, U., Fricke, K., Frommann, K., ... & Tergau, F. (2004). GABAergic modulation of DC stimulation- induced motor cortex excitability shifts in humans. *European Journal of Neuroscience*, *19*, 2720-2726.

Nitsche, M. A., & Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *The Journal of Physiology*, *527*, 633-639.

Nitsche, M. A., & Paulus, W. (2001). Sustained excitability elevations induced by transcranial DC motor cortex stimulation in humans. *Neurology*, *57*, 1899-1901.

Nitsche, M. A., & Paulus, W. (2011). Transcranial direct current stimulation—update 2011. *Restorative Neurology and Neuroscience*, *29*, 463-92.

Nitsche, M. A., Schauenburg, A., Lang, N., Liebetanz, D., Exner, C., Paulus, W., & Tergau, F. (2003b). Facilitation of implicit motor learning by weak transcranial direct current stimulation of the primary motor cortex in the human. *Journal of Cognitive Neuroscience*, *15*, 619-26.

Palm, U., Keeser, D., Schiller, C., Fintescu, Z., Nitsche, M., Reisinger, E., & Padberg, F., (2008). Skin lesions after treatment with transcranial direct current stimulation (tDCS). *Brain Stimulation*, *1*, 386-387.

Parent, A. (2004). Giovanni Aldini: From animal electricity to human brain stimulation. *The Canadian Journal of Neurological Sciences*, *31*, 576-584.

Parkin, B. L., Ekhtiari, H., & Walsh, V. F. (2015). Non-invasive human brain stimulation in cognitive neuroscience: A primer. *Neuron*, *87*, 932-945.

Paulus, W., Antal, A., & Nitsche, M. A. (2012). Physiological basis and methodological aspects of transcranial electric stimulation (tDCS, tACS, and tRNS). In C. Miniussi, W. Paulus, P. M. Rossini (Eds.), *Transcranial brain stimulation* (pp. 93-110). Boca Raton, Fl: Taylor & Francis Group.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2006). Comparison of two methods to detect publication bias in meta-analysis. *Jama, 295*, 676-680.

Podda, M. V., Cocco, S., Mastrodonato, A., Fusco, S., Leone, L., Barbati, S. A., ... & Grassi, C. (2016). Anodal transcranial direct current stimulation boosts synaptic plasticity and memory in mice via epigenetic regulation of BDNF expression. *Scientific Reports*, *6*, 22180.

Poreisz, C., Boros, K., Antal, A., & Paulus, W. (2007). Safety aspects of transcranial direct current stimulation concerning healthy subjects and patients. *Brain Research Bulletin*, *72*, 208-214.

Price, A. R., & Hamilton, R. H. (2015). A re-evaluation of the cognitive effects from single-session transcranial direct current stimulation. *Brain Stimulation, 8*, 663-665.

Price, A. R., McAdams, H., Grossman, M., & Hamilton, R. H. (2015). A meta-analysis of transcranial direct current stimulation studies examining the reliability of effects on language measures. *Brain Stimulation, 8*, 1093-1100.

Priori, A. (2003). Brain polarization in humans: a reappraisal of an old tool for prolonged non-invasive modulation of brain excitability. *Clinical Neurophysiology*, *114*, 589-595.

Priori, A., Berardelli, A., Rona, S., Accornero, N., & Manfredi, M. (1998). Polarization of the human motor cortex through the scalp. *Neuroreport*, *9*, 2257-2260.

Purpura, D. P., & McMurtry, J. G. (1965). Intracellular activities and evoked potential changes during polarization of motor cortex. *Journal of Neurophysiology, 28*, 166-185.

Ranieri, F., Podda, M. V., Riccardi, E., Frisullo, G., Dileone, M., Profice, P., ... & Grassi, C. (2012). Modulation of LTP at rat hippocampal CA3-CA1 synapses by direct current stimulation. *Journal of Neurophysiology*, *107*, 1868-1880.

Redfearn, J. W. T., Lippold, O. C. J., & Costain, R. (1964). Preliminary account of the clinical effects of polarizing the brain in certain psychiatric disorders. *The British Journal of Psychiatry*, *110*, 773-785.

Reis, J., Schambra, H. M., Cohen, L. G., Buch, E. R., Fritsch, B., Zarahn, E., ... & Krakauer, J. W. (2009). Noninvasive cortical stimulation enhances motor skill acquisition over multiple days through an effect on consolidation. *Proceedings of the National Academy of Sciences*, *106*, 1590-1595.

Rodríguez, N., Opisso, E., Pascual-Leone, Á., & Soler, M. D. (2014). Skin lesions induced by transcranial direct current stimulation (tDCS). *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation*, *7*, 765-767.

Rozisky, J. R., Antunes, L. C., Brietzke, A. P., de Sousa, A. C., & Caumo, W. (2016). Transcranial direct current stimulation and neuroplasticity. In L. Rogers (Ed.), *Transcranial direct current stimulation (tDCS): Emerging used, safety and neurobiological effects* (pp. 1-26). New York, NY: Nova Science Publishers Inc.

Russo, R., Wallace, D., Fitzgerald, P. B., & Cooper, N. R. (2013). Perception of comfort during active and sham transcranial direct current stimulation: A double blind study. *Brain Stimulation*, *6*, 946-951.

Santiesteban, I., Banissy, M. J., Catmur, C., & Bird, G. (2012). Enhancing social ability by stimulating right temporoparietal junction. *Current Biology*, *22*, 2274-2277.

Santiesteban, I., Banissy, M. J., Catmur, C., & Bird, G. (2015). Functional lateralization of temporoparietal junction–imitation inhibition, visual perspective- taking and theory of mind. *European Journal of Neuroscience*, *42*, 2527-2533.

Sellaro, R., Nitsche, M. A., & Colzato, L. S. (2016). The stimulated social brain: Effects of transcranial direct current stimulation on social cognition. *Annals of the New York Academy of Sciences*, *1369*, 218-239.

Sellaro, R., Nitsche, M. A., & Colzato, L. S. (2017). Transcranial direct current stimulation. In L. S. Colzato (Ed.), *Theory-driven approaches to cognitive enhancement* (pp. 99-112). Switzerland: Springer International Publishing.

Shiozawa, P., Fregni, F., Benseñor, I. M., Lotufo, P. A., Berlim, M. T., Daskalakis, J. Z., ... & Brunoni, A. R. (2014). Transcranial direct current stimulation for major depression: an updated systematic review and meta-analysis. *International Journal of Neuropsychopharmacology*, *17*, 1443-1452.

Silvanto, J., Muggleton, N., & Walsh, V. (2008). State-dependency in brain stimulation studies of perception and cognition. *Trends in Cognitive Sciences*, *12*, 447-454.

Stagg, C. J., Best, J. G., Stephenson, M. C., O'Shea, J., Wylezinska, M., Kincses, Z. T., ... & Johansen-Berg, H. (2009). Polarity-sensitive modulation of cortical neurotransmitters by transcranial stimulation. *Journal of Neuroscience*, *29*, 5202-5206.

Summers, J. J., Kang, N., & Cauraugh, J. H. (2016). Does transcranial direct current stimulation enhance cognitive and motor functions in the ageing brain? A systematic review and meta-analysis. *Ageing Research Reviews*, *25*, 42-54.

Utz, K. S., Dimova, V., Oppenländer, K., & Kerkhoff, G. (2010). Electrified minds: Transcranial direct current stimulation (tDCS) and galvanic vestibular stimulation

(GVS) as methods of non-invasive brain stimulation in neuropsychology—a review of current data and future implications. *Neuropsychologia*, *48*, 2789-2810.

Zaehle, T., Sandmann, P., Thorne, J. D., Jäncke, L., & Herrmann, C. S. (2011). Transcranial direct current stimulation of the prefrontal cortex modulates working memory performance: Combined behavioural and electrophysiological evidence. *BMC Neuroscience*, *12*, 2.

Chapter 5. Transcranial current stimulation of the temporoparietal junction improves lie detection

This chapter is presented as a published article and is an exact copy of the following journal publication:

Sowden, S., Wright, G. R., Banissy, M. J., Catmur, C., & Bird, G. (2015). Transcranial current stimulation of the temporoparietal junction improves lie detection. *Current Biology*, *25*, 2447-2451.

Corresponding Author:

Sophie Sowden, MRC Social, Genetic and Developmental Psychiatry, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, SE5 8AF, United Kingdom.

sophie.sowden@kcl.ac.uk

Report

Current Biology

Transcranial Current Stimulation of the Temporoparietal Junction Improves Lie Detection

Highlights

- Inconsistency between one's own and another's stated opinion impairs lie detection
- Stimulation of the right TPJ improves lie detection in opinion-inconsistent situations

Authors

Sophie Sowden, Gordon R.T. Wright, Michael J. Banissy, Caroline Catmur, Geoffrey Bird

Correspondence

sophie.sowden@kcl.ac.uk

In Brief

Sowden et al. demonstrate how a mechanism involved in inhibiting imitation and in perspective taking extends to lie detection. They show that inconsistency between the opinions of the self and another impedes lie detection and that electrical stimulation of the temporoparietal junction improves performance in these opinion-inconsistent situations.

Sowden et al., 2015, Current Biology *25*, 2447–2451
September 21, 2015
http://dx.doi.org/10.1016/j.cub.2015.08.014

CellPress

Current Biology
Report

Transcranial Current Stimulation of the Temporoparietal Junction Improves Lie Detection

Sophie Sowden,[1,*] Gordon R.T. Wright,[2] Michael J. Banissy,[3,4] Caroline Catmur,[1,5] and Geoffrey Bird[1,4]
[1]MRC Social, Genetic, and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology, and Neuroscience, King's College London, London SE5 8AF, UK
[2]Department of Psychological Sciences, Birkbeck, University of London, London WC1E 7HX, UK
[3]Department of Psychology, Goldsmiths, University of London, London SE14 6NW, UK
[4]Institute of Cognitive Neuroscience, University College London, London WC1N 3AR, UK
[5]School of Psychology, University of Surrey, Guildford GU2 7XH, UK
*Correspondence: sophie.sowden@kcl.ac.uk
http://dx.doi.org/10.1016/j.cub.2015.08.014

SUMMARY

The ability to detect deception is of vital importance in human society, playing a crucial role in communication, cooperation, and trade between societies, businesses, and individuals. However, numerous studies have shown, remarkably consistently, that we are only slightly above chance when it comes to detecting deception [1]. Here we investigate whether inconsistency between one's own opinion and the stated opinion of another impairs judgment of the veracity of that statement, in the same way that one's own mental, affective, and action states, when inconsistent, can interfere with representation of those states in another [2]. Within the context of lie detection, individuals may be less accurate when judging the veracity of another's opinion when it is inconsistent with their own opinion. Here we present a video-mediated lie-detection task to confirm this prediction: individuals correctly identified truths or lies less often when the other's expressed opinion was inconsistent with their own (experiment 1). Transcranial direct current stimulation (tDCS) of the temporoparietal junction (TPJ) has previously been shown to improve the ability to selectively represent the self or another [3–5]. We therefore predicted that TPJ stimulation would enable lie detectors to inhibit their own views, enhance those of the other, and improve their ability to determine whether another was presenting their true opinion. Experiment 2 confirmed this second prediction: anodal tDCS of the TPJ improved lie detection specifically when one's own and others' views were conflicting.

RESULTS AND DISCUSSION

Despite the frequency of deception in everyday life [6] and the importance of detecting deception within human society, humans are remarkably consistent in their inability to detect deception. Meta-analyses demonstrate a mean success rate of 54% across all published studies of lie detection ability, where chance performance is 50%, with a measurement-corrected SD of just 0.8% [1]. Although one or two cues have been demonstrated to signal deception at above-chance levels in some studies (e.g., response latency [7]), these cues are mostly overlooked in favor of non-diagnostic behavioral cues (such as the avoidance of eye contact [8]) or person-level cues relating to the perceived deceptiveness of the individual rather than what they are saying ("demeanor bias" [9]), resulting in poor performance. Here we investigate the existence of a further factor that may decrease the accuracy of lie detection when one is attempting to determine the veracity of another's stated opinion: inconsistency between one's own opinion and that of another.

It is well-established that self-representations can interfere with representation of another even when task irrelevant. The act of planning or executing an action interferes with the perception of an incongruent action performed by another [10], one's own affective state biases perception of another's incongruent affective state [11], one's own visual perspective interferes with the representation of another's spatially inconsistent visual perspective [12], and the contents of one's own mental states interfere with representation of those of another when they differ from our own [13]. A body of previous research has highlighted how each of these social abilities recruits a mechanism to enable the individual to control, or switch between, representation of the self and of others to avoid interference between inconsistent representations, such that representation of the self is enhanced and the other inhibited, or representation of the other is enhanced and the self inhibited according to task demands [14–18]. These results raise the possibility that holding an opinion inconsistent with that expressed by another may interfere with the ability to judge the veracity of the expressed opinion and that increasing the ability to inhibit representation of one's own opinion and enhance that of the other may result in improved lie-detection performance when opinions are inconsistent. Accordingly, over two experiments, participants were asked to complete a video-mediated lie-detection task based on the false-opinion paradigm [19–21] (Figure 1A), in which they were

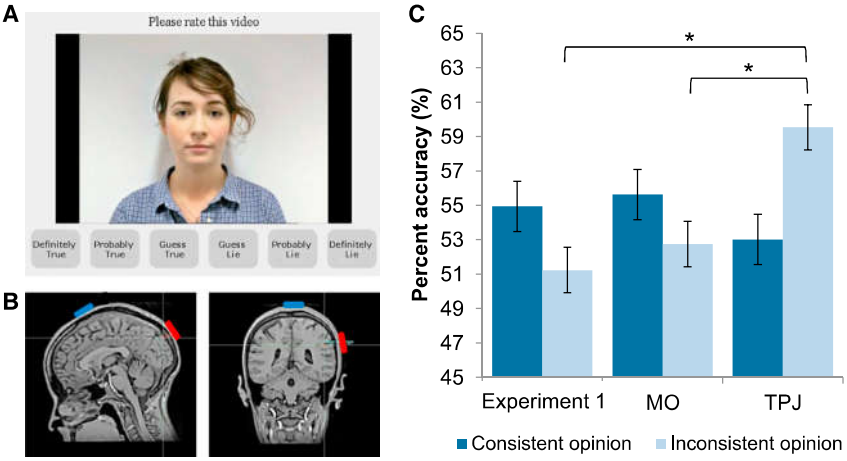

Figure 1. Example Lie-Detection Trial, Placement of Electrodes for rTPJ and MO Groups, and Visual Display of the Results

(A) Example trial from the video-mediated lie-detection task.

(B) Placement of anodal (red) and cathodal (blue) electrodes for both the MO (left) and TPJ (right) groups.

(C) Percentage accuracy when the veracity of opinion statements consistent and inconsistent with the participant's own opinion was judged. The data presented are from experiment 1 and the MO and TPJ stimulation (experiment 2) groups. * denotes significant difference at $p < 0.05$, and error bars represent the SEM.

asked to rate whether an individual (the "sender") had expressed their true or a false opinion.

Experiment 1 sought to establish evidence for the hypothesized opinion inconsistency effect by comparing lie-detection performance for opinion-consistent (where the participant's opinion matched the sender's stated opinion) and opinion-inconsistent (where the participant's opinion was opposite to the sender's stated opinion) statements. A group of healthy adult volunteers (n = 63; mean age = 33.5, SD = 6.4; 44 female) were asked to complete the lie-detection task after completing a questionnaire ascertaining their views on a number of controversial topics. As hypothesized, when rating the veracity of opinion statements expressed by senders (i.e., whether the sender had presented their true opinion), participants were significantly more accurate when the view expressed by the sender was consistent with their own view (mean percent accuracy ± SEM: 54.9% ± 0.8%) than when inconsistent (51.2% ± 1.1%; $t(62) = 3.02$, $p = 0.004$, $d = 0.49$).

Experiment 2 tested the efficacy of anodal transcranial direct current stimulation (tDCS), a form of non-invasive electrical brain stimulation, of the temporoparietal junction (TPJ) to improve lie-detection accuracy on opinion-inconsistent trials. A series of recent studies suggest that the TPJ plays a crucial role in the mechanism that enables the control of representations of the self and of others [2–5, 15, 18]. These studies suggest the TPJ allows representation of the self to be inhibited and the other enhanced, or the self enhanced and the other inhibited, and that this process is recruited in theory of mind, imitation inhibition, and visual perspective taking [17, 18, 22–25]. In line with previous demonstrations across these different social domains, it was predicted that stimulation of the TPJ would allow representation of one's own opinion to be inhibited and representation of the other's opinion to be enhanced, leading to a reduction in the opinion inconsistency effect and improved lie-detection performance on opinion-inconsistent trials.

Thirty-three healthy adult participants (mean age = 24.2, SD = 4.6; 18 female) underwent 20 min of tDCS, over either the right TPJ (rTPJ) or a mid-occipital (MO) control region (Figure 1B), prior to completing the same video-mediated lie-detection task but after their own opinions were obtained. The two groups did not differ in their age, gender, or opinions on the topics discussed ($p > 0.05$), and there was no significant difference in overall lie-detection performance between the two groups ($t(31) = 0.93$, $p = 0.361$). Moreover, overall performance (collapsing across opinion-consistent and -inconsistent trials) did not differ significantly from the population derived average of 54% [1] in either the TPJ ($t(15) = 1.18$, $p = 0.257$) or MO ($t(16) = 0.25$, $p = 0.804$) group.

Lie-detection performance was then analyzed using a mixed-effect two-way ANOVA with a within-subjects factor of opinion consistency (two levels: performance on opinion-consistent versus opinion-inconsistent trials) and a between-subjects factor of stimulation group (two levels: stimulation of TPJ versus MO). As predicted, there was a significant opinion consistency × stimulation group interaction ($F(1,31) = 7.74$, $p = 0.009$, $\eta_P^2 = 0.20$), whereby participants who underwent rTPJ stimulation were significantly more accurate when the sender's expressed opinion was inconsistent with their own opinion (59.5% ± 1.6%) when compared to those administered MO control stimulation (52.7% ± 1.2%; $t(31) = 3.32$, $p = 0.002$, $d = 1.15$). Conversely, there was no significant difference in lie-detection performance between stimulation groups during trials in which the sender and participant's opinions were consistent ($t(31) = 1.03$, $p = 0.313$; Figure 1C).

The significant improvement in lie detection after tDCS of the TPJ, specific to situations in which one must suppress one's own opinion to judge the veracity of another's statement, supports both the involvement of a self-other control mechanism in lie detection and the involvement of the TPJ in this process. Results suggest that in situations of conflict between one's own opinion and that of another, one must inhibit the representation of one's own opinion and enhance that of the other in order to successfully discriminate between a truth and a lie.

The relatively modest improvement in absolute accuracy observed after TPJ stimulation is in accordance with both the size of the consistency effect observed in experiment 1 and the degree of individual differences in the population as a whole. A mean success rate of 54% across all published studies (albeit with an unknown proportion of opinion-consistent and -inconsistent trials), with a measurement-corrected SD of just 0.8% [1], means that the increase in lie detection ability after TPJ stimulation is therefore not trivial with respect to population-level

statistics. Indeed, the mean percentage accuracy of the TPJ stimulation group on opinion-inconsistent trials was significantly greater than the population-derived figure of 54% ($t(15) = 3.33$, $p = 0.005$, $d = 1.17$) but on consistent trials was not ($t(15) = 0.46$, $p = 0.651$). Accuracy in the MO stimulation group did not differ significantly from 54% on either consistent or inconsistent trials ($p > 0.05$).

These data are promising for understanding the mechanisms involved in, and factors affecting, lie-detection performance. However, care should be taken when considering the use of tDCS to improve lie detection in real-world situations, where even the 87% lie-detection accuracy achieved by polygraphs [26] does not ensure that truths are not flagged as lies and vice versa—especially in situations where the base rate of lying is likely to be low, such as in employment screening. Also, although the time course of the improvement in lie detection was not measured in this study, it is likely to be short lived: with the stimulation parameters used in the current experiment, one would only expect effects to last for approximately 90 min [27]. Thus, it seems that much of the value of these results lies in their theoretical implications for the processes involved in lie detection, for the applicability of self-other control mechanisms to higher-level social cognition, and for our knowledge of TPJ function.

The specificity of the improvement in lie detection (on opinion-inconsistent trials only) suggests that stimulation does not act on a general process involved in social cognition, such as theory of mind [28] or emotion recognition [29]. Rather, the results are best explained in the context of the mechanism of self-other control. Self-other control has been shown to underlie the role of the TPJ in various social functions on the basis of organic lesions [25], experimentally induced disruption [30], and facilitation [3–5] of TPJ function and through the use of neuroimaging methods such as fMRI [14].

With regard to the full mechanism by which opinion inconsistency produces impaired lie-detection performance, it is plausible that individuals who hold strong opinions for or against a topic have been exposed to similar arguments in support of their position and against the opposite view [31], allowing them to recognize the match between a position and the most common justifications for that position when a sender has an opinion consistent with their own. If individuals are presented with an opinion counter to their own, their lack of experience relating to justifications supporting that stance on the topic means they may be ill equipped to judge the appropriateness of the justification and therefore the veracity of the stated opinion. It is important to realize, however, that the effect of TPJ stimulation is unlikely to be attributable wholly to inhibition of one's own opinion: if inhibition were the only effect of stimulation, then performance on inconsistent trials would have been equivalent, rather than superior, to performance on consistent trials. Instead, results suggest that representation of the other was also enhanced, in line with previous findings on self-other control and TPJ stimulation [3–5] and with theories attributing an attentional, switching, or gating role to the TPJ [16, 32, 33]. A necessary subsequent focus for research is to identify the mechanism by which an enhanced representation of the other results in an increased ability to detect lies.

In conclusion, experiment 1 established self-other interference effects in the context of lie detection, where representations refer to the opinions of the self and the other. Participants were less accurate at distinguishing truthful from false opinions when the sender's opinion was inconsistent with their own. Experiment 2 demonstrated that this performance interference effect could be reduced through anodal stimulation of TPJ, improving lie detection specifically on those trials in which this effect was most prominent. These results suggest that boosting the ability to control representations of the self and other—in this case inhibiting one's own opinion in order to more accurately represent that of the other—can improve lie detection in opinion-inconsistent situations.

EXPERIMENTAL PROCEDURES

Lie-Detection Task

In both experiments 1 and 2, participants began by completing an "opinion questionnaire" in which they gave their opinion on 20 topics. For each item they rated the degree to which they agreed or disagreed with the topic on a six-point scale (with an answer of "1" demonstrating strong agreement, and "6" strong disagreement, with a topic). Example items on the questionnaire include "euthanasia," "medical abortion," "genetically modified foodstuffs," and "animal testing." During the lie-detection task, participants watched a series of 80 randomly ordered video clips of individuals ("senders") expressing their views, as well as a brief justification of their view, on the same topics included in the opinion questionnaire. These took the form of "I am in favor of euthanasia because everyone deserves a chance to die with dignity." After watching each video, participants were asked to rate whether the sender had presented their true opinion or whether they had lied, on a 6-point scale (1, definitely true; 6, definitely lie; see Figure 1A). The task took a total of 25 min.

The stimulus set comprised the same set of 40 truthful and 40 deceptive statements for all participants, conveyed by 20 different individuals (ten males and ten females). The stimulus set contained four video statements about each of the 20 topics contained in the opinion questionnaire—two truths and two lies—and in a fully factorial design, two statements were spoken in agreement and two in disagreement with each topic. The videos were recorded during a previous experiment and were all provided by individuals who had strong opinions for or against each topic (ratings of "1" or "6" on the opinion questionnaire). Note that the factorial combination of truthful and deceptive statements, for and against each topic, means that the observed improvement in lie detection performance on opinion-inconsistent trials after stimulation is an effect on accuracy rather than bias in both experiment 1 and experiment 2.

Data Analysis

Trials were divided into opinion-consistent and opinion-inconsistent trials on a trial-by-trial basis according to the sender's expressed opinion and the participant's opinion as reported on the opinion questionnaire. For example, if a participant listed their opinion as "against" a topic, a trial in which the sender expressed their opinion as "against" the topic was classed as an opinion-*consistent* trial, whereas a trial in which the sender expressed their opinion as "for" the topic was classed as an opinion-*inconsistent* trial. Participants' responses on each trial were dichotomized as either a "truth" (reponses 1–3) or a "lie" (responses 4–6) judgment to account for individual differences in the use of the extremities of the rating scale. The percentage accuracy of judgments constituted the measure of lie detection performance, which was compared in experiment 1 for opinion-consistent and -inconsistent trials using a paired-sample t test. In experiment 2, lie-detection performance was analyzed using a mixed-effect two-way ANOVA (with opinion consistency as the within-subjects factor and stimulation group as the between-subjects factor). Two participants were excluded prior to data analysis in experiment 2: one from the TPJ stimulation group, who responded "true" to over 90% of trials, and one from the MO stimulation group, who completed less than 20% of trials, leaving too few trials for analysis.

It should be noted that the design of both experiment 1 and experiment 2 followed current best practice guidelines by comparing lie detection performance on the critical condition of interest (opinion-inconsistent trials) with a

within-participant baseline (opinion-consistent trials), which also served as an extremely closely matched control condition. Indeed, participants were performing the same task, to videotaped statements from the same people, concerning the same topics, on both opinion-consistent and opinion-inconsistent trials.

Experiment 2 tDCS Protocol

Participants were randomly assigned to one of two tDCS groups: rTPJ (n = 17) or MO control (n = 18). All participants were healthy volunteers, with no known developmental or neurological disorders, normal or corrected-to-normal vision, and no contraindications to tDCS. Prior to study completion, all participants were naive to the aims of the experiment.

All participants underwent anodal stimulation, induced with two saline-soaked surface sponge electrodes (35 cm^2) and delivered by a battery-driven, constant current stimulator. According to group assignment, the anodal electrode was placed over central parietal 6 (CP6) for rTPJ stimulation and occipital zero (OZ) for MO control stimulation (electroencephalography 10/20 system). MO was used as an active control site as there was no a priori reason to assume that stimulation to this region would differ from baseline. The cathodal electrode was placed over the vertex as a reference point, individually measured on each participant (Figure 1B). A weak electrical current (1 mA) was delivered offline (preceding the task) for a total of 20 min, following the procedure used by Santiesteban and colleagues [4, 5], as these effects are reported to be more robust than online stimulation. The effects of stimulation with these parameters have been demonstrated to last for 90 min after stimulation [27]. Participants completed the opinion questionnaire prior to stimulation, whereas the lie detection task was completed within the critical 90 min window after stimulation. Both experiments 1 and 2 received full ethical approval by the local Research Ethics Committees (Birkbeck, University of London and Institute of Psychiatry, Psychology, and Neuroscience, King's College London, respectively).

ACKNOWLEDGMENTS

S.S is supported by a doctoral studentship from the Medical Research Council. M.J.B is supported by the Economic and Social Research Council (ES/K00882X/1). We would like to thank Henry Woodward and Katerina Capouskova for their assistance with the testing of the participants in experiment 2.

Received: June 19, 2015
Revised: July 23, 2015
Accepted: August 7, 2015
Published: September 3, 2015

REFERENCES

1. Bond, C.F., and Depaulo, B.M. (2008). Individual differences in judging deception: accuracy and bias. Psychol. Bull. *134*, 477–492.
2. Decety, J., and Sommerville, J.A. (2003). Shared representations between self and other: a social cognitive neuroscience view. Trends Cogn. Sci. *7*, 527–533.
3. Hogeveen, J., Obhi, S.S., Banissy, M.J., Santiesteban, I., Press, C., Catmur, C., and Bird, G. (2015). Task-dependent and distinct roles of the temporoparietal junction and inferior frontal cortex in the control of imitation. Soc. Cogn. Affect. Neurosci. *10*, 1003–1009.
4. Santiesteban, I., Banissy, M.J., Catmur, C., and Bird, G. (2012). Enhancing social ability by stimulating right temporoparietal junction. Curr. Biol. *22*, 2274–2277.
5. Santiesteban, I., Banissy, M.J., Catmur, C., and Bird, G. Functional lateralization of temporoparietal junction: Imitation inhibition, visual perspective taking and theory of mind. Eur. J. Neurosci. Published online August 3, 2015. http://dx.doi.org/10.1111/ejn.13036.
6. The Global Deception Research Team (2006). A world of lies. J. Cross Cult. Psychol. *37*, 60–74.
7. Seymour, T.L., and Kerlin, J.R. (2008). Successful detection of verbal and visual concealed knowledge using an rt-based paradigm. Appl. Cogn. Psychol. *490*, 475–490.
8. Mann, S., Vrij, A., Leal, S., Granhag, P.A., Warmelink, L., and Forrester, D. (2012). Windows to the soul? Deliberate eye contact as a cue to deceit. J. Nonverbal Behav. *36*, 205–215.
9. Levine, T.R., Serota, K.B., Shulman, H., Clare, D.D., Park, H.S., Shaw, A.S., Shim, J.C., and Lee, J.H. (2011). Sender demeanor: individual differences in sender believability have a powerful impact on deception detection judgments. Hum. Commun. Res. *37*, 377–403.
10. Miall, R.C., Stanley, J., Todhunter, S., Levick, C., Lindo, S., and Miall, J.D. (2006). Performing hand actions assists the visual discrimination of similar hand postures. Neuropsychologia *44*, 966–976.
11. Silani, G., Lamm, C., Ruff, C.C., and Singer, T. (2013). Right supramarginal gyrus is crucial to overcome emotional egocentricity bias in social judgments. J. Neurosci. *33*, 15466–15476.
12. Keysar, B., Barr, D.J., Balin, J.A., and Brauner, J.S. (2000). Taking perspective in conversation: the role of mutual knowledge in comprehension. Psychol. Sci. *11*, 32–38.
13. Happé, F.G.E. (1994). An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. J. Autism Dev. Disord. *24*, 129–154.
14. Brass, M., Derrfuss, J., and von Cramon, D.Y. (2005). The inhibition of imitative and overlearned responses: a functional double dissociation. Neuropsychologia *43*, 89–98.
15. Brass, M., Ruby, P., and Spengler, S. (2009). Inhibition of imitative behaviour and social cognition. Philos. Trans. R. Soc. Lond. B Biol. Sci. *364*, 2359–2367.
16. Cook, J.L. (2014). Task-relevance dependent gradients in medial prefrontal and temporoparietal cortices suggest solutions to paradoxes concerning self/other control. Neurosci. Biobehav. Rev. *42*, 298–302.
17. Santiesteban, I., White, S., Cook, J., Gilbert, S.J., Heyes, C., and Bird, G. (2012). Training social cognition: from imitation to Theory of Mind. Cognition *122*, 228–235.
18. Spengler, S., von Cramon, D.Y., and Brass, M. (2009). Control of shared representations relies on key processes involved in mental state attribution. Hum. Brain Mapp. *30*, 3704–3718.
19. Mehrabian, A. (1971). Nonverbal betrayal of feeling. J. Exp. Res. Pers. *5*, 64–73.
20. Wright, G.R., Berry, C.J., and Bird, G. (2012). "You can't kid a kidder": association between production and detection of deception in an interactive deception task. Front. Hum. Neurosci. *6*, 87.
21. Wright, G.R., Berry, C.J., and Bird, G. (2013). Deceptively simple … The "deception-general" ability and the need to put the liar under the spotlight. Front. Neurosci. *7*, 152.
22. Decety, J., and Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. Neuroscientist *13*, 580–593.
23. Sowden, S., and Shah, P. (2014). Self-other control: a candidate mechanism for social cognitive function. Front. Hum. Neurosci. *8*, 789.
24. Spengler, S., Bird, G., and Brass, M. (2010). Hyperimitation of actions is related to reduced understanding of others' minds in autism spectrum conditions. Biol. Psychiatry *68*, 1148–1155.
25. Spengler, S., von Cramon, D.Y., and Brass, M. (2010). Resisting motor mimicry: control of imitation involves processes central to social cognition in patients with frontal and temporo-parietal lesions. Soc. Neurosci. *5*, 401–416.
26. Gougler, M., Nelson, R., Handler, M., Krapohl, D., Shaw, P., and Bierman, L. (2011). Meta-analytic survey of criterion accuracy of validated polygraph techniques. Polygraph *40*, 194–305.
27. Nitsche, M.A., and Paulus, W. (2001). Sustained excitability elevations induced by transcranial DC motor cortex stimulation in humans. Neurology *57*, 1899–1901.
28. Saxe, R., and Young, L. (2013). Theory of Mind: How brains think about thoughts. In The Oxford Handbook of Cognitive Neuroscience,

Volume 2, K. Ochsner, and S.M. Kosslyn, eds. (Oxford University Press), pp. 204–213.

29. Tseng, L.Y., Tseng, P., Liang, W.K., Hung, D.L., Tzeng, O.J., Muggleton, N.G., and Juan, C.H. (2014). The role of superior temporal sulcus in the control of irrelevant emotional face processing: A transcranial direct current stimulation study. Neuropsychologia *64C*, 124–133.

30. Sowden, S., and Catmur, C. (2015). The role of the right temporoparietal junction in the control of imitation. Cereb. Cortex *25*, 1107–1113.

31. Nickerson, R.S. (1998). Confirmation bias: a ubiquitous phenomenon in many guises. Rev. Gen. Psychol. *2*, 175.

32. Bzdok, D., Langner, R., Schilbach, L., Jakobs, O., Roski, C., Caspers, S., Laird, A.R., Fox, P.T., Zilles, K., and Eickhoff, S.B. (2013). Characterization of the temporo-parietal junction by combining data-driven parcellation, complementary connectivity analyses, and functional decoding. Neuroimage *81*, 381–392.

33. Mitchell, J.P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. Cereb. Cortex *18*, 262–271.

Chapter 6. The impact of autistic traits on the ability to control competing representations of self and other's opinions.

As discussed in Chapter 2 of the current thesis, Autism Spectrum Disorder (ASD) is increasingly considered as a disorder of top-down control or modulation of social behaviour (Cook, Barbalat, & Blakemore, 2012; Wang & Hamilton, 2012). This chapter considers self-other control as a candidate mechanism to explain impairments in social behaviour in these individuals. A group of neurotypical individuals and a group of individuals with a clinical diagnosis of ASD completed the newly developed lie detection task described in Chapter 5 (Sowden, Wright, Banissy, Catmur, & Bird, 2015) as an index of self-other control, probing its relationship with autistic trait severity.

6.1. Introduction

6.1.1. A theory for the social deficits in ASD. Despite early support for the theory that a dysfunctional mirror neuron system may account for the broad social deficits observed in ASD (Iacoboni & Dapretto, 2006; Williams et al., 2006), there is now an increasingly large body of research which does not support this theory (Hamilton, 2013; Southgate & Hamilton, 2008; Sowden, Koehne, Catmur, Dziobek, & Bird, 2016a). Under the 'broken mirror' account, automatic imitation, as a behavioural index of mirror neuron system function, should be reduced in ASD (Williams, Whiten, & Singh, 2004; Williams, Whiten, Suddendorf, & Perrett, 2001). However, a number of studies report intact, and even hyper-imitative behaviour in these individuals (Bird, Leighton, Press, & Heyes, 2007; Grossi, Marcone, Cinquegrana, & Gallucci, 2012; Schunke et al., 2016; Sowden, et al., 2016a; Spengler, Bird, & Brass, 2010a). This is consistent with clinical symptoms of echolalia and echopraxia (the involuntary copying of the speech and actions of others) in ASD (Grossi et al., 2012).

Consequently, an alternative theory was proposed, in line with reports of enhanced imitative behaviour in ASD as well as impairments in various other domains key for social functioning. This states that individuals with autism may be compromised in their ability to *control* or *modulate* their social behaviour (Cook et al., 2012; Sowden & Shah, 2014; Wang & Hamilton, 2012). For example, individuals with ASD fail to show the typical modulatory effects of social priming on their imitative behaviour (Cook & Bird, 2012), whilst also showing high correlations between hyperimitation and deficits in theory of mind (Spengler et al., 2010a). Social impairments are correlated across domains in ASD (Spengler et al., 2010a; Spengler, von Cramon, & Brass, 2009, 2010b), and a deficiency within a neural network which supports the top-down control of representations of 'self' and 'other' seems plausible to account for such a behavioural profile (Brass, Derrfuss, & von Cramon, 2005; Brass, Ruby, & Spengler, 2009; Sowden & Shah, 2014).

Such 'self-other control' is characterised as the ability to control, or switch between, competing representations held for the self and for others with whom one interacts. This is fundamental to many aspects of social cognition. For example, to successfully take another person's perspective, representation of one's own perspective must be inhibited, and representation of the other person's perspective enhanced. Similarly, to pass a theory of mind task one must represent the beliefs, desires, or intentions of another person whilst inhibiting one's own beliefs, desires, and intentions. Likewise, when empathising with another person, the affective state that results from representation of the other's emotions must be distinguished from one's own (Singer & Lamm 2009). Finally, the control of imitation (also referred to as imitation-inhibition) requires one to switch between motor representations activated by the observation of another's action and self-generated motor representations (Brass et al. 2009). This mechanism of self-other control is able to explain correlated performance across tasks in which self- and other-representations are likely to be co-activated

and must be selected between (Spengler et al., 2009, 2010a, 2010b), as well as behavioural training induced transfer effects (de Guzman, Bird, Banissy, & Catmur, 2016; Santiesteban et al., 2012a) between these tasks, and brain stimulation induced improvement across tasks which require inhibition and enhancement of self or other representations (Hogeveen et al., 2015; Santiesteban, Banissy, Catmur, & Bird, 2012b; 2015).

Given this body of evidence, it is possible that impaired self-other control may explain the reduced top-down modulation of behaviour commonly observed in ASD. This idea was first introduced in Chapter 2 of this thesis, which demonstrated that poorer self-other control (as indexed by a decreased ability to control imitative response tendencies and thus increased imitation) is associated with increased symptom severity in individuals with ASD.

The present study, however, aimed to move beyond the use of the 'imitation-inhibition' or 'control of imitation' task to investigate whether an impairment of self-other control can explain performance in other social domains in ASD. Accordingly, a task was developed which could elicit self-other interference effects and be amenable to modulation via non-invasive brain stimulation, in a similar nature to the imitation-inhibition task. Rather than manipulating the congruence of self- and other-related motoric representations, this task manipulates the congruency of the *opinions* of the self and others in the context of lie detection. The task allows lie detection performance to be assessed under situations of opinion consistency as well as inconsistency between the self and others. See Chapter 5 (Sowden et al., 2015) for a full outline of the theoretical development and validation of this task as a suitable measure of self-other control.

6.1.2. Control of the opinions of self and other. It is consistently reported that humans perform only slightly better than chance in detecting lies, with accuracy reported at

54% (with a measurement corrected SD of 0.8%) across all experimental lie detection studies (Bond & Depaulo, 2008). There has been little success in identifying factors which might influence this ability. However, Sowden et al. (2015), utilising a newly developed video-mediated lie detection task, presented one factor which appears to modulate one's ability to judge the veracity of other individuals' opinion statements: (in)consistency between the opinion of the self and others.

In a similar way to the imitation-inhibition task – where inconsistency between one's own motor plan and that of the other interferes with performance (Brass, Bekkering, Wohlschläger, & Prinz, 2000) – inconsistency between the opinions of the self and others was found to interfere with veracity judgements, decreasing accuracy in such situations to 51%. When opinions were consistent between the self and the other accuracy was at 55%. Thus, it appears that self-other interference effects, observed across a range of social cognitive abilities (Brass et al., 2009; Santiesteban et al., 2012a), are also present in the context of lie detection, whereby self-other control may be required to overcome a conflict of opinion and allow successful performance. For example, if an individual expresses an opposing opinion to your own regarding euthanasia, one must inhibit representation of one's own opinion on euthanasia and enhance representation of the opinion being expressed by the other person in order to successfully judge the veracity of their opinion statement. Interestingly, work by Stel and colleagues (Stel, va Dijk, & Olivier, 2009) ratifies the link between lie detection performance and the control of competing representations of the self and other. Individuals with a greater ability to inhibit imitation (i.e. improved self-other control) performed better when assessing the veracity of emotional expressions made by another individual.

Notably, Sowden et al. (2015) provide further support for the video-mediated lie detection task being considered as a task to index self-other control, with the use of non-

invasive transcranial direct current stimulation (tDCS). This study demonstrated an improvement in lie detection performance selectively under situations of self-other opinion inconsistency after utilising the same protocol of excitatory tDCS to the right temporoparietal junction as used to investigate self-other control in imitation-inhibition, perspective-taking and theory of mind (Hogeveen et al., 2015; Santiesteban et al., 2012b, 2015).

Therefore, the current study sought to extend investigation of the integrity of self-other control in ASD by considering lie detection; a social domain requiring the online representation and control of competing representations of self and other. It was predicted that both individuals with ASD and a matched Control group would show significant consistency effects (higher veracity judgement accuracy during opinion-consistent than opinion-inconsistent situations). A group difference in the size of these consistency effects was also hypothesised, whereby individuals with ASD, in line with the hypothesis of reduced self-other control in this population, may show larger consistency effects than their neurotypical counterparts, as well as increasing consistency effects with increasing autistic trait severity across the whole sample.

6.2. Materials and Methods

6.2.1. Participants. Twenty-one high-functioning individuals with a clinical diagnosis of ASD (4 female; mean age = 35.2 years, SD = 11.2) and an age, gender and IQ-matched sample of 39 healthy Control individuals (10 female; mean age = 35.2 years, SD = 12.2) were recruited from a database held at the Institute of Psychiatry, King's College London. All participants reported normal or corrected-to-normal vision. All individuals in the ASD group had an existing, independent, clinical diagnosis of autism or ASD according to DSM-IV criteria (American Psychiatric Association, 1994), which was confirmed with the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000).

Autism Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001) scores were significantly higher [$t(58) = 11.7$, $p < .001$] in the ASD group (mean = 37.1, SD = 7.5) than Control group (mean = 16.3, SD = 6.0), and the groups did not differ significantly in age [$t(58) = 0.01$, $p = .990$], proportion of females [$\chi^2(1) = .33$, $p = .565$], or IQ [$t(58) = 0.11$, p = .911]. Full sample characteristics are presented in Table 1.

Table 1

Participant Characteristics.

	ASD	**Control**
N	21	39
Gender	17 Male, 4 Female	29 Male, 10 Female
Mean Age (years)	35.2 (*11.2*)	35.2 (*12.2*)
Mean Full-scale IQ	109.9 (*16.0*)	110.3 (*13.0*)
Mean AQ	37.1 (*7.5*)	16.3 (*6.0*)
ADOS Classification	11 Autism, 10 Autism Spectrum	n/a
Mean ADOS-G Score	10.2 (*2.3*)	n/a

NB. Values in brackets represent standard deviation from the mean.

6.2.2. Experimental procedure. Participants completed the Opinion Questionnaire (Sowden et al., 2015), AQ and Toronto Alexithymia Scale (TAS-20; Bagby, Parker, & Taylor, 1994), followed by a video-mediated lie detection task (as outlined in Chapter 5 of this thesis; Sowden et al., 2015) as a behavioural measure of self-other control concerning the opinions of the self and others.

6.2.2.1. Opinion Questionnaire. The Opinion Questionnaire (see Appendix A) requires the participant to rate the degree to which they are for or against a number of topics on a 6-point scale (with an answer of '1' demonstrating they are strongly for, and '6' strongly

against, the topic). A total of 20 topics are asked about, with example items on the questionnaire including ‘medical abortion’, ‘euthanasia’, ‘stem cell research’, and ‘animal testing’. Responses for each item were dichotomised for later data analysis as either ‘for’ (responses 1-3) or ‘against’ (responses 4-6). This dichotomisation was completed to account for individual differences in the use of the extremities of the rating scale.

6.2.2.2. Autism Spectrum Quotient. All participants completed the AQ (see Appendix B), a 50-item self-report questionnaire designed to assess severity of autistic-like traits both in the atypical and general population. Although the AQ was not designed as a diagnostic tool, it is frequently used to screen typical participants for autistic traits (Hurst, Mitchell, Kimbrel, Kwapil, & Nelson-Gray, 2007). An overall score indicative of the level of ‘autistic-like traits’ was calculated in this study for use in data analysis.

6.2.2.3. Toronto Alexithymia Scale. The Toronto Alexithymia Scale (TAS-20; see Appendix C) is a 20-item self-report measure of an individual’s ability to identify, describe and experience one’s own emotions. Reponses are made on a 5-point Likert scale, from 1 (‘does not describe me well’) to 5 (‘describes me very well’). An overall score for the level of ‘alexithymic traits’ was derived from this questionnaire for use in data analysis.

6.2.2.4. Lie detection task. The lie detection task involved watching a series of 80 randomly-ordered video clips of individuals (‘Senders’) expressing their views, as well as a brief justification for their view, on the same topics included in the Opinion Questionnaire. Example dialogue from a trial is “I am in favour of euthanasia because everyone deserves a chance to die with dignity.” Following each video, participants rated whether they thought the Sender had presented their true opinion, or had lied about their opinion, on a 6-point scale (Definitely True – Definitely Lie; see Figure 1). The task took a total of 25 minutes.

The stimulus set comprised 40 truthful and 40 deceptive statements, conveyed by 20 different individuals (10 males and 10 females). 4 video statements were observed for each of the 20 topics contained in the Opinion Questionnaire: 2 truths and 2 lies; and in a fully-factorial design, 2 statements were spoken in agreement and 2 in disagreement with each topic. The videos were recorded during a previous experiment and were all provided by individuals who had strong opinions for or against each topic (ratings of '1' or '6' on the Opinion Questionnaire).

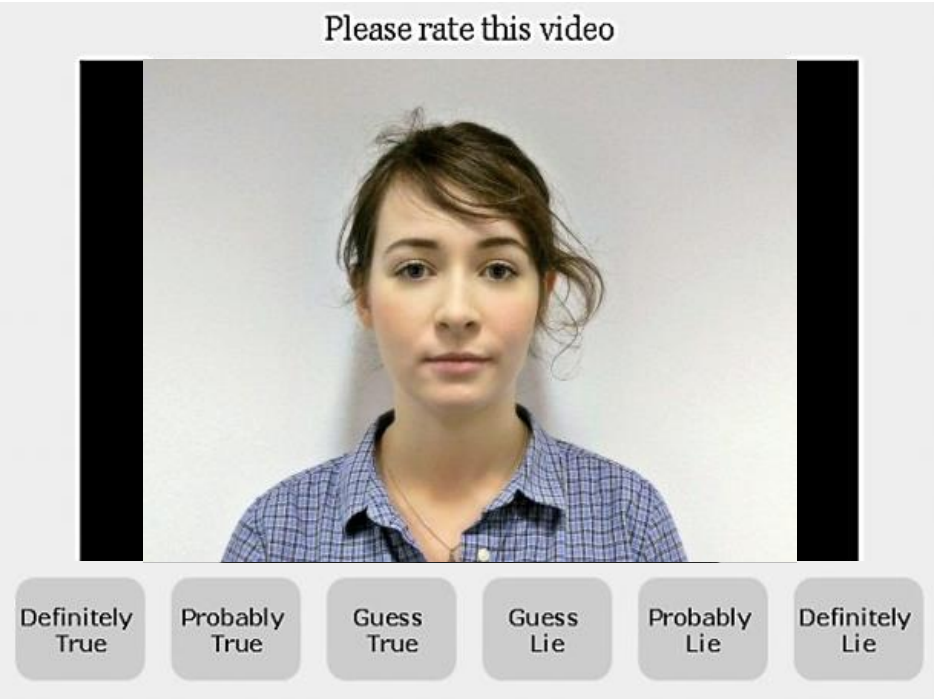

*Figure 1***.** Example of on-screen display during the video-mediated lie detection task, showing a Sender presenting their opinion on a topic and the scale with which participants are required to rate the veracity of each video statement heard. Full permission was gained for the use of this image.

6.3. Results

Trials were divided into opinion-consistent and opinion-inconsistent trials on a trial-by-trial basis according to the Sender's expressed opinion, and the participant's opinion as reported on the Opinion Questionnaire. For example, if a participant expressed agreement

with a topic, a trial in which the Sender also expressed agreement with the topic was classed as an opinion-*consistent* trial, whereas a trial in which the Sender expressed their disagreement with the topic was classed as an opinion-*inconsistent* trial. Participants' responses on each trial were dichotomised as either a 'truth' (responses 1-3) or a 'lie' judgment (responses 4-6), again to account for individual differences in the use of the extremities of the rating scale. Percentage accuracy of judgments constituted the measure of lie detection performance and an individual's opinion consistency effect, as a specific type of self-other interference effect, was calculated as lie detection performance accuracy on opinion-consistent minus opinion-inconsistent trials, with smaller consistency effects indicating improved self-other control (i.e less impact on lie detection performance due to opinion inconsistency).

Data (shown in Figure 2) were analysed using a mixed effect two-way analysis of variance (ANOVA), with opinion consistency as the within-subjects factor and group (ASD or Control) as the between-subjects factor. These analyses revealed a significant group*consistency interaction under our one-tailed hypothesis ($F(1,58) = 3.03$, $p = .044$, $\eta_p^2 = .05$). The main effect of opinion consistency approached, but did not reach, significance under our one-tailed hypothesis ($F(1,58) = 2.37$, $p = .065$, $\eta_p^2 = .04$). Examining the consistency effects separately in the two groups revealed a trend towards a significant consistency effect in the ASD group when using the one-tailed significance value ($t(20) = 1.71$, $p = .052$, $d = 0.66$), whereby accuracy was increased during opinion-consistent trials (mean = 57.3%, standard error of the mean [SEM] = 1.7%) relative to opinion-inconsistent trials (mean = 51.9%, SEM = 1.8%). The failure to achieve standard significance levels was perhaps unsurprising given that the experiment was underpowered to detect an effect of size similar to that found in Sowden et al., (2015), with power calculations indicating a required sample size of 40 participants. Data collection was restricted to the number of individuals

with ASD available to take part during the data collection period. However, a medium effect size was observed here. A consistency effect was not observed in the Control group at a one-tailed significance value ($t(38) = 0.192$, $p = .424$; opinion-consistent performance mean = 53.7%, SEM = 1.2%, opinion-inconsistent performance mean = 54.0%, SEM = 1.4%).

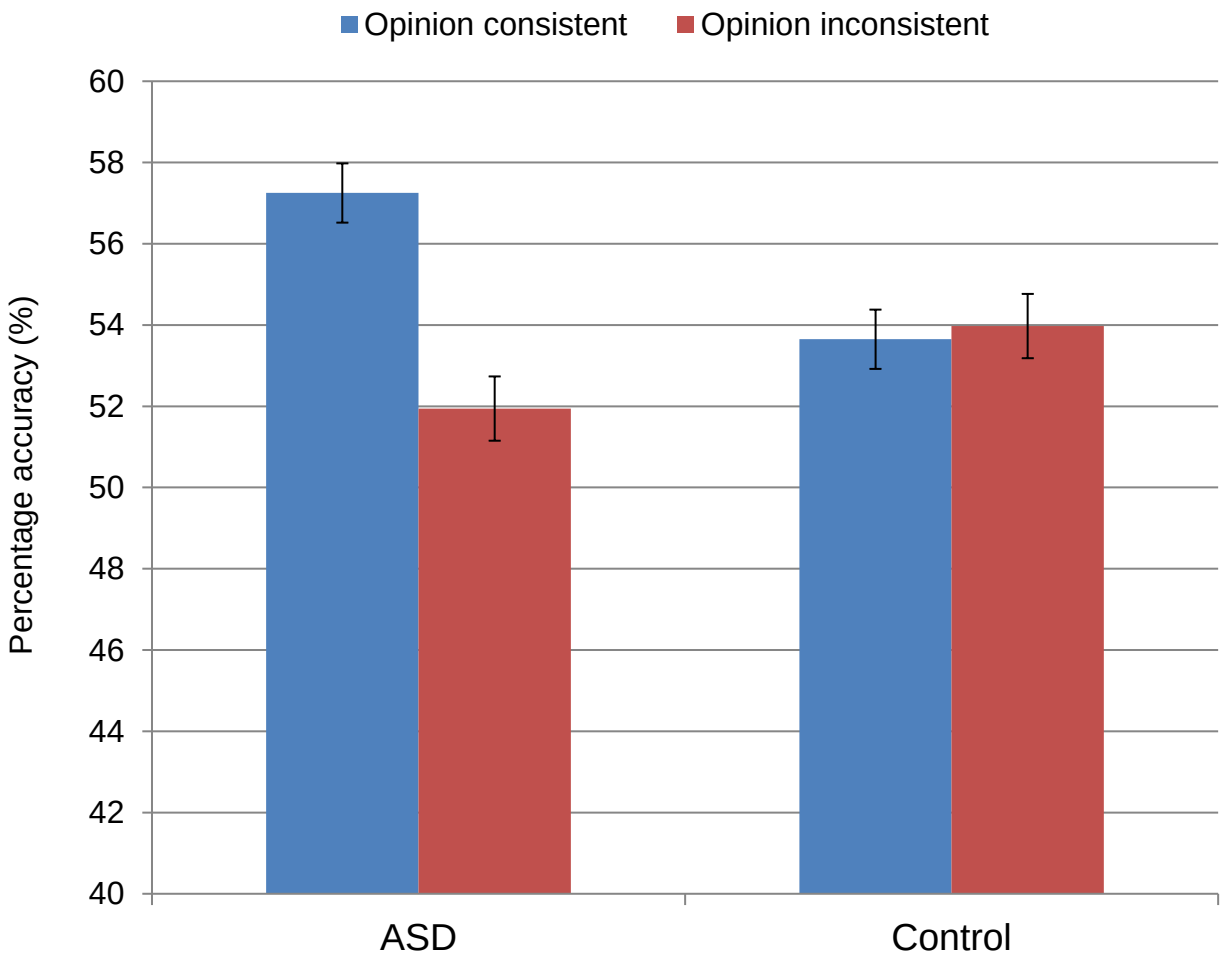

Figure 2. Percentage accuracy when the veracity of opinion statements consistent and inconsistent with the participant's own opinion was judged. The data presented are from the ASD and Control groups separately and error bars represent the standard error of the mean.

The presence of a significant group*consistency interaction in the ANOVA merits further investigation of the relationship between autistic traits and opinion consistency effects. Simple bivariate correlations revealed a significant positive correlation (see Figure

3a) between lie detection opinion consistency (consistent – inconsistent trial performance accuracy) and autistic traits, as measured by the AQ (r = .433, p = .001). Interestingly, this relationship was observed to be significant in both the ASD (r = .436, p = .048) and Control (r = .488, p = .002) groups independently (Figure 3b and 3c). No relationship was found between ADOS and AQ scores or between ADOS severity scores and opinion consistency in the ASD group (ps > .05).

Consistent with the well-established relationship between alexithymia and autism/autistic-like traits (Aaron et al., 2015), AQ and TAS-20 scores were highly correlated in the current sample (r = .701, p < .001). Moreover, there is now a body of research to suggest that both alexithymia (Bird & Cook, 2013; Quattrocki & Friston, 2014; Sowden, Brewer, Catmur, & Bird, 2016b) and demographic variables such as age, gender and IQ (Harms et al., 2010) are associated with performance in social cognitive tasks. Thus, regression analyses were conducted to test whether autistic traits account for variance over and above that explained by these other factors.

A hierarchical regression was conducted to model the variance in the size of the opinion consistency effect. Age, gender, IQ and TAS-20 scores were entered into the first step of the regression model. The first level model of the regression was not significant (p = .196) and neither were any of the variables included independent predictors of opinion consistency effects (ps > .05). AQ scores were entered into the second step of the model and revealed AQ scores to be a significant predictor of opinion consistency effects (β = 0.69, $t(58) = 3.91$, p < .001), over and above the other variables included in the model. The addition of AQ scores significantly improved the fit of the model, increasing the variance accounted for by 23.6% ($F(1,54) = 15.32$, p < .001), compared to 3.7% accounted for by the first step of the model.

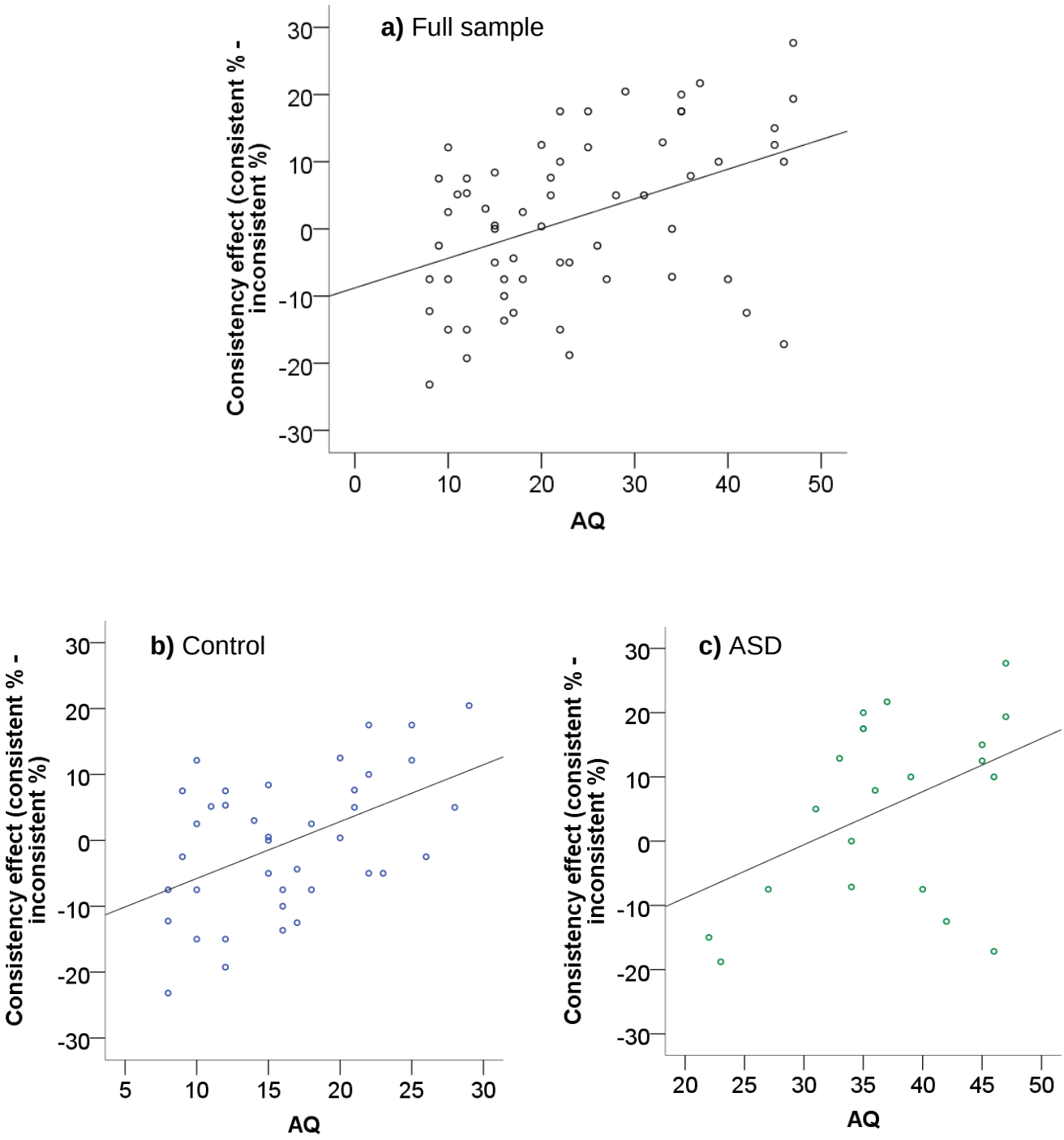

Figure 3. Scatterplots of the correlations between opinion consistency effects (opinion-consistent percent accuracy - opinion-inconsistent percent accuracy) and AQ scores in **(a)** the full sample, **(b)** the Control group and **(c)** the ASD group.

6.4. Discussion

The current study sought to investigate the integrity of self-other control in ASD in the context of lie detection as a newly-developed index of self-other control. This was achieved by comparing lie detection performance under situations of consistency and inconsistency between the opinions of the self and others. It was theorised that impaired self-other control would lead to a greater opinion consistency effect, whereby lie detection accuracy would decrease to a greater extent with opinion inconsistency. Such self-other interference effects in the context of opinions and lie detection have been found previously (Sowden et al., 2015), and are thought to be analogous to those observed in other social domains such as imitation-inhibition, perspective-taking and empathy (de Guzman et al., 2016; Santiesteban et al., 2012a).

Despite a trend towards a significant consistency effect in the predicted direction in the ASD group, whereby performance accuracy was higher in opinion-consistent relative to inconsistent trials, there was an absence of a consistency effect in the Control group. Thus, this study presents a failure to replicate the original consistency effect as found by Sowden et al. (2015). It should be noted however, that the current study was underpowered to detect effect sizes of the magnitude reported in Sowden et al. (2015). In contrast, and in line with predictions, the consistency effect was significantly associated with autistic traits across the whole spectrum of autistic traits; as AQ scores increased, so did the consistency effect in the video-mediated lie detection task. This was found in the current study in two independent samples; one sample of individuals with a clinical diagnosis of ASD and one sample of healthy Control individuals.

In conclusion, a number of factors support further larger scale investigation of self-other control as a candidate mechanism to explain performance across social domains in

ASD. There is a need to move towards a more dimensional approach in our investigation of cognition and behavioural functioning in autism (Landa & Goldberg, 2005) and indeed performance deficits are shown to correlate across social domains (Spengler et al., 2009, 2010a, 2010b) as well as correlating with autism trait severity (Sowden et al., 2016a). Moreover, despite the lack of an overall consistency effect in the present Control group, a significant and reliable relationship was identified here between AQ scores and the opinion consistency effect. Together, this body of evidence provides a starting point for considering the therapeutic use of domain-general training in self-other control in individuals with ASD and perhaps even across other disorders of social cognition such as schizophrenia which are associated with a similar pattern of impairments.

6.5. References

Aaron, R. V., Benson, T. L., & Park, S. (2015). Investigating the role of alexithymia on the empathic deficits found in schizotypy and autism spectrum traits. *Personality and Individual Differences, 77*, 215-220.

American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders (3th ed.). Washington, DC: American Psychiatric Association Press.

Bagby, R. M., Taylor, G. J., & Parker, J. D. (1994). The twenty-item Toronto Alexithymia Scale—II. Convergent, discriminant, and concurrent validity. *Journal of Psychosomatic Research, 38*, 33–40.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31*, 5–17.

Bond, C. F., & DePaulo, B. M. (2008). Individual differences in judging deception: Accuracy and bias. *Psychological Bulletin, 134*, 477-492.

Brass, M., Bekkering, H., Wohlschläger, A., & Prinz, W. (2000). Compatibility between observed and executed finger movements: Comparing symbolic, spatial, and imitative cues. *Brain and Cognition, 44*, 124-143.

Brass, M., Derrfuss, J., & von Cramon, D. Y. (2005). The inhibition of imitative and overlearned responses: A functional double dissociation. *Neuropsychologia, 43*, 89-98.

Brass, M., Ruby, P., & Spengler, S. (2009). Inhibition of imitative behaviour and social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*, 2359-2367.

Bird, G., & Cook, R. (2013). Mixed emotions: The contribution of alexithymia to the emotional symptoms of autism. *Translational Psychiatry, 3*, e285.

Bird, G., Leighton, J., Press, C., & Heyes, C. (2007). Intact automatic imitation of human and robot actions in autism spectrum disorders. *Proceedings of the Royal Society B: Biological Sciences, 274*, 3027-3031.

Cook, J., Barbalat, G., & Blakemore, S.J. (2012). Top-down modulation of the perception of other people in schizophrenia and autism. *Frontiers in Human Neuroscience, 6*, 175.

Cook, J. L., & Bird, G. (2012). Atypical social modulation of imitation in autism spectrum conditions. *Journal of Autism and Developmental Disorders, 42*, 1045-1051.

de Guzman, M., Bird, G., Banissy, M. J., & Catmur, C. (2016). Self–other control processes in social cognition: from imitation to empathy. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*, 20150079**.**

Grossi, D., Marcone, R., Cinquegrana, T., & Gallucci, M. (2012). On the differential nature of induced and incidental echolalia in autism. *Journal of Intellectual Disability Research, 57*, 903-912.

Hamilton, A. F. D. C. (2013). Reflecting on the mirror neuron system in autism: A systematic review of current theories. *Developmental Cognitive Neuroscience, 3*, 91-105.

Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies. *Neuropsychology Review, 20*, 290-322.

Hogeveen, J., Obhi, S. S., Banissy, M. J., Santiesteban, I., Press, C., Catmur, C., & Bird, G. (2015). Task-dependent and distinct roles of the temporoparietal junction and inferior frontal cortex in the control of imitation. *Social Cognitive and Affective Neuroscience*, *10*, 1003-1009.

Hurst, R. M., Mitchell, J. T., Kimbrel, N. A., Kwapil, T. K., & Nelson-Gray, R. O. (2007). Examination of the reliability and factor structure of the Autism Spectrum Quotient (AQ) in a non-clinical sample. *Personality and Individual Differences*, *43*, 1938-1949.

Iacoboni, M., & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience*, *7*, 942-951.

Landa, R. J., & Goldberg, M. C. (2005). Language, social, and executive functions in high functioning autism: A continuum of performance. *Journal of Autism and Developmental Disorders*, *35*, 557-573.

Lord, C., Risi, S., Lambrecht, L., Cook Jr, E.H., Leventhal, B.L., DiLavore, P.C.,. . . & Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, *30*, 205–223.

Quattrocki, E., & Friston, K. (2014). Autism, oxytocin and interoception. *Neuroscience and Biobehavioral Reviews*, *47*, 410–430.

Santiesteban, I., Banissy, M. J., Catmur, C., & Bird, G. (2012b). Enhancing social ability by stimulating right temporoparietal junction. *Current Biology*, *22*, 2274-2277.

Santiesteban, I., Banissy, M. J., Catmur, C., & Bird, G. (2015). Functional lateralization of temporoparietal junction–imitation inhibition, visual perspective- taking and theory of mind. *European Journal of Neuroscience*, *42*, 2527-2533.

Santiesteban, I., White, S., Cook, J., Gilbert, S. J., Heyes, C., & Bird, G. (2012a). Training social cognition: From imitation to theory of mind. *Cognition, 122*, 228-235.

Schunke, O., Schöttle, D., Vettorazzi, E., Brandt, V., Kahl, U., Bäumer, T., ... & Brass, M. (2016). Mirror me: Imitative responses in adults with autism. *Autism, 20*, 134-144.

Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences, 1156*, 81-96.

Southgate, V., & Hamilton, A. F. D. C. (2008). Unbroken mirrors: Challenging a theory of autism. *Trends in Cognitive Sciences, 12*, 225-229.

Sowden, S., Brewer, R., Catmur, C., & Bird, G. (2016b). The specificity of the link between alexithymia, interoception, and imitation. *Journal of Experimental Psychology. Human Perception and Performance, 42*, 1687-1692.

Sowden, S., Koehne, S., Catmur, C., Dziobek, I., & Bird, G. (2016a). Intact automatic imitation and typical spatial compatibility in autism spectrum disorder: Challenging the broken mirror theory. *Autism Research, 9*, 292-300.

Sowden, S., & Shah, P. (2014). Self-other control: A candidate mechanism for social cognitive function. *Frontiers in Human Neuroscience, 8*, 789.

Sowden, S., Wright, G. R., Banissy, M. J., Catmur, C., & Bird, G. (2015). Transcranial current stimulation of the temporoparietal junction improves lie detection. *Current Biology, 25*, 2447-2451.

Spengler, S., Bird, G., & Brass, M. (2010a). Hyperimitation of actions is related to reduced understanding of others' minds in autism spectrum conditions. *Biological Psychiatry, 68*, 1148-1155.

Spengler, S., von Cramon, D. Y., & Brass, M. (2009). Control of shared representations relies on key processes involved in mental state attribution. *Human Brain Mapping, 30*, 3704-3718.

Spengler, S., von Cramon, D.Y., & Brass, M. (2010b). Resisting motor mimicry: control of imitation involves processes central to social cognition in patients with frontal and temporo-parietal lesions. *Social Neuroscience, 5*, 401–416.

Stel, M., van Dijk, E., & Olivier, E. (2009). You want to know the truth? Then don't mimic!. *Psychological Science*, *20*, 693-699.

Wang, Y., & Hamilton, A.F.D.C. (2012). Social top-down response modulation (STORM): A model of the control of mimicry in social interaction. *Frontiers in Human Neuroscience*, *6*, 153.

Williams, J. H., Waiter, G. D., Gilchrist, A., Perrett, D. I., Murray, A. D., & Whiten, A. (2006). Neural mechanisms of imitation and 'mirror neuron' functioning in autistic spectrum disorder. *Neuropsychologia*, *44*, 610-621.

Williams, J. H., Whiten, A., & Singh, T. (2004). A systematic review of action imitation in autistic spectrum disorder. *Journal of Autism and Developmental Disorders*, *34*, 285-299.

Williams, J. H., Whiten, A., Suddendorf, T., & Perrett, D. I. (2001). Imitation, mirror neurons and autism. *Neuroscience & Biobehavioral Reviews*, *25*, 287-295.

Chapter 7. Quantifying compliance and acceptance through public and private social conformity

Social conformity refers to a general class of social influence whereby exposure to the attitudes and beliefs of a group causes an individual to alter their own attitudes and behaviour to be more similar to those of the group. Thus, we see another candidate domain for self-other performance interference effects in the context of higher-level group social interaction. This chapter describes the development and pilot of a novel task to measure two varieties of social influence (acceptance and compliance); one which occurs when the participant is required to respond in private and one when responding publicly to a group of others. The chapter is written from a methodological standpoint, with a focus on dissociating two varieties of social conformity, and thus two levels of self-other interference effects, within the same task. However, with respect to the current thesis, supplementary data show how such self-other interference effects relate to population level autistic traits, as a primer for future research using this task to further understand the neural and behavioural bases of self-other control.

7.1. Introduction

The multitude of ways in which the behaviour and attitudes of others impact our own has been studied since the very inception of psychology (Cialdini & Goldstein, 2004). A particular focus has been the study of how groups influence the behaviour of the individual, with studies such as those of Asch on social conformity (Asch, 1951, 1955, 1956), some of the most well-known in the field. In these studies, participants were asked to complete a simple perceptual task (judging the length of lines) in a group setting where, unbeknown to the participant, the other members of the group were confederates of the experimenter. During critical trials, despite the task having an obvious answer, the confederates all gave the

incorrect answer. Only a quarter of participants remained completely independent of the majority, with the rest showing various degrees of conformity towards the group's responses.

Subsequent work has identified various types of group influence, individuated by factors including the circumstances of the influence (e.g. whether the group pressure is explicit or implicit), and the nature of the change brought about in the individual. With respect to the latter, of interest to the current study is the distinction between compliance and acceptance (Kelman, 1958; Nail, Di Domenico, & MacDonald, 2013). Compliance and acceptance can be distinguished based on the type of attitude change brought about by the social influence. Compliance occurs when the individual publicly agrees with the group but does not change their own attitude or belief, whereas acceptance occurs when the social influence causes the individual to internalise the belief or attitude expressed by the group such that it becomes their own. Compliance and acceptance are thought to arise primarily from normative and informational influence, respectively (Abrams, Wetherell, Cochrane, Hogg, & Turner, 1990; Deutsch & Gerard, 1955). Normative influence occurs due to the desire of individuals to be accepted by the group, or at least not to be in conflict with the group. Abrams et al. (1990) suggest that "compliance with the demands and expectations of other group members and overt agreement with their views occurs because of their power to reward, punish, accept or reject individual members." In contrast, informational influence is thought to result in acceptance because it occurs when individuals look to others for evidence as to the state of the world. As such, its effects are thought to be maximal when the state of the world is ambiguous, or when the individual is uncertain about a decision or judgement (Cialdini & Goldstein, 2004; Cialdini & Trost, 1998; Deutsch & Gerard, 1955).

Although it is commonly accepted that both types of social influence are typical in everyday social situations, social conformity effects obtained using the Asch paradigm are usually attributed to normative influence (compliance) only (e.g. Allen, 1965, 1975; Bond &

Smith, 1996; Turner, 1991). The fact that the perceptual decision task has an obviously correct answer (participants almost never make errors during trials where confederates give the correct answer) is usually taken as *prima facie* evidence that results occur due to compliance to the group decision through normative influence. This view is not universally accepted however (e.g. Abrams et al., 1990; Turner, 1985), and claims of an informational influence are supported by a handful of studies that have compared levels of conformity using this task between groups of individuals who must respond publicly, and those who have the opportunity to make their responses in private.

The logic of these experiments is that, by comparing individuals who give their responses in public with those who respond in private, the relative contributions of normative and informational influence (and hence compliance versus acceptance) can be established. Individuals who respond in private should experience little to no normative influence due to the fact that the group members are unaware if they have conformed or not, and thus any group influence should be due to informational influence alone. Comparison of the degree of social conformity in the private and public groups therefore allows the existence of normative and informational influence to be established using the standard Asch paradigm.

Results of studies which have compared public and private conformity effects support the existence of an informational effect in the Asch paradigm as well as a normative effect. For example, Asch (1956; Experiment 4) found that rates of conformity (the percentage of critical trials across all participants in which errors in the direction of the confederates' judgements were made) dropped from 43% in public conditions to 12.5% in private conditions – demonstrating a substantial normative influence effect. However, the 12.5% conformity rate in the private conformity condition was higher than the 1% error rate observed in control groups who were not subject to group pressure to give incorrect answers. This indicates the presence of an informational influence albeit of smaller magnitude than

that of the normative influence. Abrams et al. (1990) reached a similar conclusion, noting that participants conformed on an average of 58% of trials when asked to respond publicly, but only 33% when responding privately (see also Deutsch & Gerard, 1955, who reported a 16% drop in conformity using a privacy manipulation).

While it is logically coherent to compare public and private responses to identify the relative contribution of normative and informational influence in the Asch paradigm, the current implementation of this comparison can be improved. Thus far, the manipulation of public and private responses has been between groups: the public group make their responses as per the original paradigm, whereas in the private group, although the confederates give their responses publicly, the participant responds in private. Groups of participants who respond publicly are then compared with those who respond in private. One issue with this approach is that comparisons between groups have the potential to be affected by sampling error and do not account for subject-level variance; this means that between-subject comparisons are always less sensitive than within-subjects comparisons. This loss of sensitivity is especially important when using techniques with a low signal-to-noise ratio such as functional magnetic resonance imaging (fMRI), which have been used several times to examine the neural correlates of conformity-related processes (Berns et al., 2005; Campbell-Meiklejohn et al., 2012a; Klucharev, Hytönen, Rijpkema, Smidts, & Fernández, 2009; Klucharev, Munneke, Smidts, & Fernández, 2011). Perhaps more important is the potential confounding effect of asking participants to respond in private while the confederates, who the participant believes to be other participants, respond publicly. Abrams and colleagues have argued that the distinction between the way in which the group of confederates respond (publicly), and the way in which the participant responds (privately), may result in the participant feeling like an out-group member. This may cause them to anti-conform to the confederates, reducing the observed magnitude of any conformity effect due to informational

influence. Ideally then, the private / public response manipulation would be on a within-subjects level and the participants and confederates would respond in a similar manner.

The Asch task itself has been criticised on a number of methodological grounds, several of which were noted soon after the Asch experiments were originally published (e.g. Crutchfield, 1955). Chief among these criticisms is the fact that the original Asch task is an insensitive measure; the choice of only three response options, two of which are very obviously wrong, presumably means that a great deal of pressure to conform must be experienced before an incorrect option is chosen. Small effects of social influence are therefore likely to go undetected. The use of only three response options also means the size of any conformity effect is impossible to measure; therefore, the degree of conformity in the task refers to the frequency of conformity across trials, rather than the size of the effect on any one trial. The use of a continuous response scale would alleviate these problems, although as far as we are aware continuous response scales have only been used with 'off-line' conformity paradigms where participants do not interact directly with group members in real life, but instead receive false feedback as to the responses of a group who had previously completed the task (e.g. Campbell-Meiklejohn et al., 2012a, 2012b; Klucharev et al., 2009, 2011; Zaki, Schirmer, & Mitchell, 2011). Meta-analytic work has demonstrated that off-line conformity paradigms result in reduced conformity effects compared to on-line paradigms in which participants interact with group members, and off-line paradigms are thought to induce conformity through different mechanisms to on-line paradigms (Bond & Smith, 1996; Deutsch & Gerard, 1955; Levy, 1960).

This paper therefore reports preliminary data obtained using a novel on-line social conformity task based on the original Asch paradigm which was designed to measure the effects of normative and informational influence on a within-subjects level utilising public and private responses – thus identifying whether both compliance and acceptance may be

induced in individuals by the group. Furthermore, participant and confederate responses were obtained using the same method, eliminating a factor which may have resulted in the participant feeling like an out-group member and anti-conforming from group responses in previous studies. Finally, participants were able to use a continuous response scale, meaning that small effects of social influence could be detected.

Participants were asked to take part in a study on 'the effect of motor preparation on perception.' They were told that they would be tested in groups of four, for efficiency, and asked to judge the colour of a patch on a central screen. They were informed that they would need to type their responses on some trials, and speak their responses on other trials as the experiment was designed to compare the effect of preparing a manual versus a vocal response on colour perception. The order of responding was fixed such that the participant was the last to give a response. Thus, randomly across trials, it was either the case that the participant and confederates all typed their responses (baseline trials in which no social influence was present), the participant typed their response after the confederates had spoken their responses out loud (private conformity condition), the participant and confederates gave spoken responses (public conformity condition), or the participant and two of the three confederates gave a spoken response while one confederate typed their response ('reduced majority' trials in which the participant responded publicly but where the majority consisted of 2 rather than 3 confederates). Reduced majority trials were also included to ensure the response style for confederates and participants did not differ across trials (i.e. all participants experienced private trials). Given previous results it was predicted that an informational influence (indicated by private conformity) as well as a normative influence (indicated by greater public compared to private conformity) would be identified. It was also predicted that the informational effect would be smaller in magnitude than that of normative influence and, that the normative effect would be greater in magnitude as the size of the majority increases.

The presence of reduced majority trials, where only 2 confederate responses are available (compared to 3 on a standard public trial) allows for this comparison to be made.

7.2. Materials and Methods

7.2.1. Participants. Twenty-two healthy adult female individuals (mean age = 21.2 years, SD = 2.8) were recruited through the King's College London research recruitment website. Therefore, all participants were either staff or students at King's College London, from a wide variety of disciplines. For recruitment to the study, it was a requirement that participants had not previously studied psychology at college or higher education level. Only female participants were recruited in order to remove the potential for out-group effects based on the sex of the participant compared to that of the (female) confederates. Participants were informed that this was a study investigating the effect of 'motor preparation on perception' and that the investigators were interested in better understanding how motor preparation for spoken versus typed responses impacts on visual perception. They were also told that to speed up data collection, and depending on the number of participants signed up to the timeslot, they would be tested in a group of up to 4 participants. In fact, each participant performed the task alongside 3 other individuals, all of whom were confederates of the experimenter. On arrival participants reported normal or corrected to normal vision, and they were asked to confirm their area of study, or department in which they worked at King's College London. Following the experiment, participants were fully debriefed as to the true nature of the experiment. During a funnelled debrief – whereby participants were asked a series of questions which started broad and open-ended and funnelled to more specific questions to gauge their level of awareness of the true study aims – it was apparent that only one participant believed their fellow participants in the experiment to be confederates (and this one additional participant was excluded, leaving 21 participants' data for analysis).

7.2.2. Experimental procedure

7.2.2.1. Room and confederate setup. The same 3 female confederates performed the task alongside each participant. They were undergraduate psychology students from King's College London (mean age = 20.3 years), and their mean age did not differ significantly from the mean age of the participant sample ($p > .05$). In each instance, Participant 1 (confederate) arrived 15 minutes prior to the testing slot to ensure they were always first to arrive, taking a seat at Position 1 (see Figure 1 for seating position labels). Following this, the real participant would be allowed time to arrive (most arriving a few minutes early for the testing session). The experimenter would instruct the participant to enter the testing room and take a seat at any laptop. To reduce uncontrolled interaction between the participant and confederates prior to the task, they were told to sit quietly whilst reading the instruction sheet and consent form. Due to the layout of the room, every participant without prompting chose to sit at Position 4. With the testing room door open to the half way point, Position 2 was blocked for the participant to take a seat in this position, and of the two positions remaining, seating themselves at Position 3 would block another participant's access to Position 4. Thus, all participants followed the same pattern of seating themselves at Position 4 without direct instruction from the experimenter. Finally, in each instance, Participant 2 (confederate) would arrive a minute or two after the participant and take a seat at Position 2, followed by Participant 3 (confederate) who would arrive shortly after calling to ask for directions to the testing room and take the last remaining seat (Position 3). The participant number assigned to the confederates was counterbalanced across testing sessions. Finally, the 4 testing laptops (all 15.6-inch, ASUS-Z550C, running Windows 10) were labelled with a sticker denoting the participant number prior to the testing session. See Figure 1 for positions corresponding to the participant number assigned during the experiment.

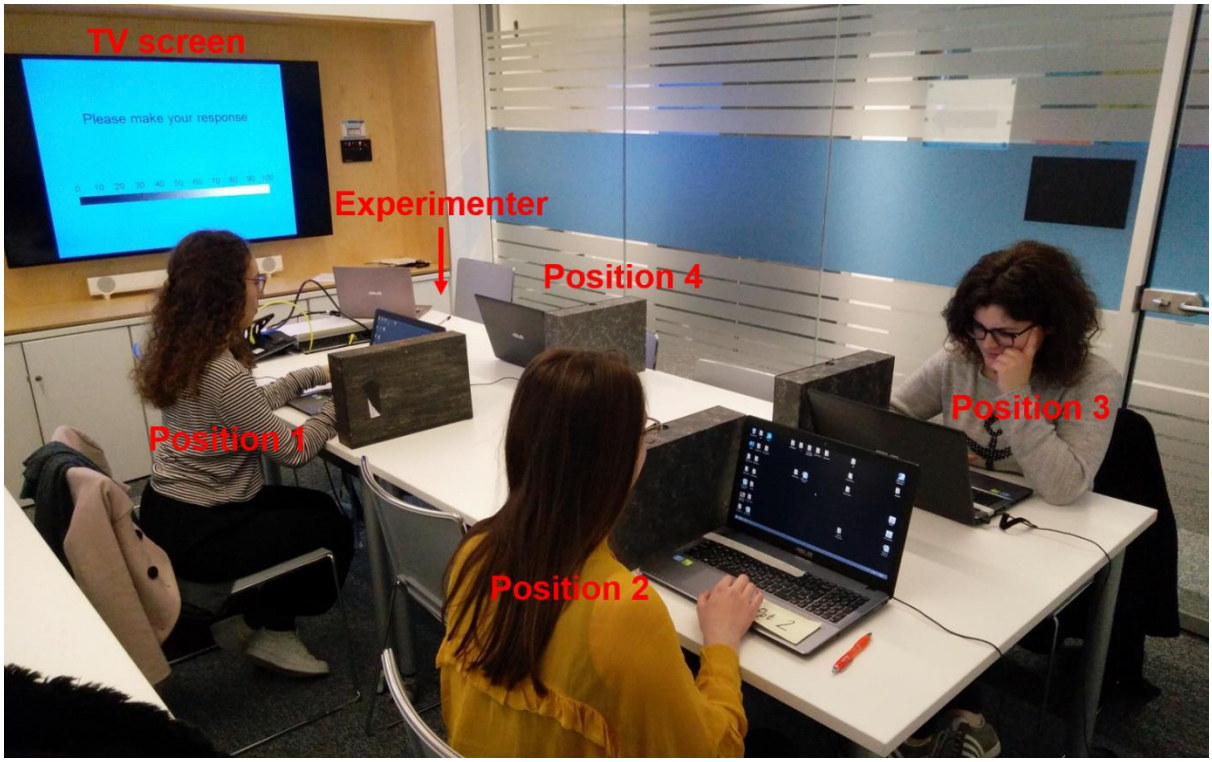

Figure 1. Room layout for the experiment showing locations of the main TV screen with respect to the experimenter and each of the 4 participant seating positions. All individuals appearing in the photo have given full permission for the use of this image.

7.2.2.2. Visual perception task. The visual perception task was programmed and run using Cogent and Cogent Graphics for Matlab. In the task, participants were required to make judgements about the colour of squares presented on a large central TV screen (Samsung-DM65D, 65-inch display, running Windows 10 software; see Figure 1 for location in the testing room). The colour of the squares could lie anywhere on a scale from white to black. Following the presentation of the square, a colour bar was presented on the TV screen. Participants were required to give a numerical response, matching the colour of the square with the same colour on the colour bar and reporting the colour's numerical value (see Figure 2). On some trials participants were asked to *type* their response using the numeric keyboard and during others they were asked to *say* their response out loud for the experimenter to

record. These instructions were given to each participant on their own laptop screen for each trial. They were instructed to pay very close attention to whether they should say or type their response on each trial, as this would sometimes differ across participants. For example, during any one trial, all participants could be required to *say* their responses out loud or they may all be required to *type* their response. Alternatively, on some trials one of the participants may be instructed to *type* their response whilst all other participants were instructed to *say* their response. Ostensibly, in order for the experimenter to record the spoken responses by hand, participants were instructed to respond in ascending participant number order.

7.2.2.3. Single trial structure. The experimenter verbally introduced each new trial and the trial number was displayed (e.g. 'Trial 1', 'Trial 2', 'Trial 3') on the main TV screen and each participant's laptop screen. Each new trial (see Figure 2 for a visual depiction of one trial) began with the trial number as well as a coloured square being presented on the main TV screen for 3 seconds. The trial number was also displayed on the participants' laptop screens during these 3 seconds. The square was then replaced by a colour bar on the main TV screen (showing shades from black to white; with labels in 10-step increments from 0-100), and simultaneously, each participant was instructed on their own laptop screen whether they should *type* or *say* their answer on that trial. Participant 1 then made their response, followed by the other participants in ascending participant order. During trials where any participant was instructed to type their response, their laptop would beep after two digits had been pressed using the numeric keyboard. This acted as an auditory cue for the next participant to respond. Participants were instructed that they could respond using any number on the colour bar scale between 01 and 99. At the end of the trial, all participants and the experimenter pressed the space bar to move the experiment onto the next trial. The protocol began with 10 practice trials, followed by 153 trials in the main experiment, with 2 short breaks after each 51 trials. The task took approximately 45 minutes to complete.

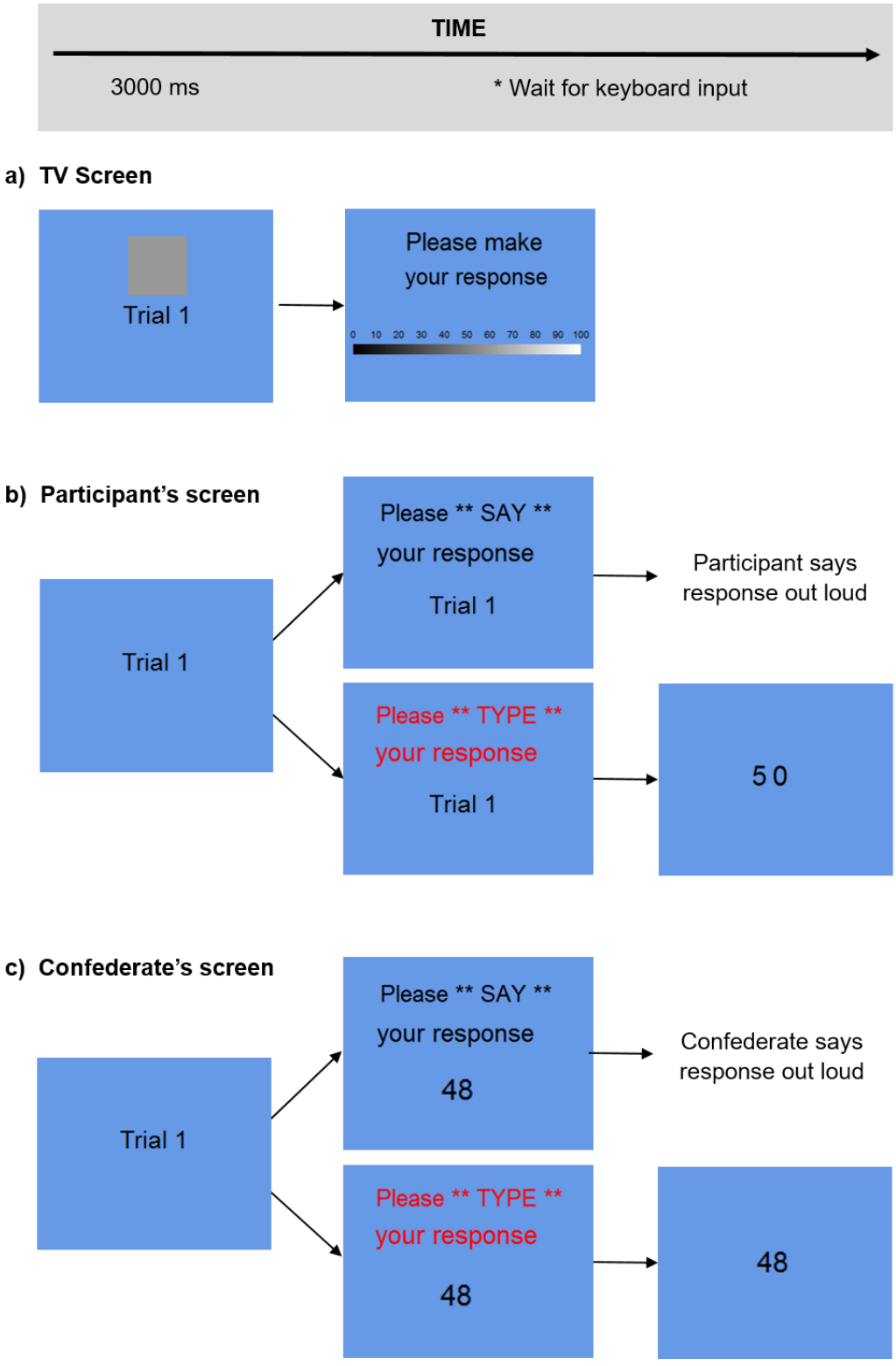

Figure 2. Example structure of one trial in the experiment showing display presented on **a)** the main TV screen (visible to all participants), **b)** the real participant's screen and **c)** a confederate's screen. * The TV screen programme waits for a space bar press to trigger the next trial, as do participant and confederate laptops during spoken trials. During typed trials, the programme waits for two digits to be typed before displaying the number on the screen, after which the programme waits for a space bar press to trigger the next trial.

7.2.2.4. Trial type manipulations. To allow for public and private performance to be assessed within the same task, different trial types were introduced. First, to gauge the participants' baseline performance when judging the colour of the squares, there were 18 'silent' trials during which all participants were instructed to *type* their response. Second, 'private' trials were introduced (27 in total) during which the participant was required to *type* their response (and therefore this response remained private) whilst all other participants were instructed to *say* their response. All 4 participants (i.e. the real participant plus each of the three confederates) experienced their own set of 27 private trials where they were required to *type* their response. These trials constitute the private condition when the participant responded privately, and reduced majority trials when each of the confederates responded privately. Finally, during 'public' trials, all participants were instructed to *say* their response (27 in total). Although these trials constitute the public condition, for analysis purposes it should be noted that during reduced majority trials the participant was required to their response publicly after two responses from confederates, therefore these trials could be considered public trials with only two confederate responses rather than three. Based on the seminal studies of Asch (1951, 1955), one may expect to also see a conformity effect on reduced majority trials, although of lesser magnitude than on the public trials with three confederate responses. Across the whole experiment, responses on 30% of trials for each participant were typed, and 70% spoken.

7.2.2.5. Confederate congruency manipulation. In line with Asch's original conformity studies (Asch, 1951, 1955, 1956), within each of the private and public conditions there was a ratio of 1:2 congruent to incongruent trials. Here, congruency refers to the relationship between the correct response and the responses given by the confederates. During congruent trials confederates gave a correct answer, while during incongruent trials they gave an incorrect answer. Confederates were instructed as to which response to give via

their laptop screens. During each trial, according to the trial type, a number was calculated for each confederate to be instructed to respond by taking the correct square colour (e.g. 50) and adding or subtracting values from this. On an incongruent trial, 15 was added to or subtracted from this value (e.g. 35 or 65), as well as a small amount of jitter between 0 and 3 (e.g. values between 32 and 38 on a trial where the confederates responded lower than the correct response of 50, or between 62 and 68 on a trial where the confederates responded higher than the correct response of 50). For congruent trials, however, only the jitter of 0, 1, 2, or 3 was added to or subtracted from the correct response (e.g. confederate responses could vary between 47 and 53 for a trial where the correct response was 50).

7.3. Results

For each trial a 'response discrepancy' was calculated as the difference between the participant's response and the correct response. These values were calculated relative to the direction of the responses of the confederates. Thus, positive values indicate a discrepancy in the participant's response in the direction of the confederates' responses and negative values indicate a discrepancy in the opposite direction to that of the confederates. For example, for a trial where the correct response was 50 and confederates responded higher on the scale, a response of 55 by the participant would be considered a response discrepancy of +5, whereas a response of 45 would be considered a response discrepancy of -5. However, for a trial where the correct response was again 50 but confederates responded lower on the scale, a response of 55 would be a response discrepancy of -5, whilst a response of 45 would be a response discrepancy of +5. This allows for a calculation of the degree to which participants shift their responses towards or away from the confederates' responses during private and public trials. The private and public conformity effects were calculated as the difference between response discrepancies during their respective congruent and incongruent trials, thus providing a measure of the informational influence effect (indexed by the size of the private

effect) and the additional normative influence during public trials (indexed by the size of the public effect minus the private effect).

Trials for which response discrepancies were ±2.5 standard deviations from each participant's mean for each condition were discarded with an a priori threshold of 15% lost trials for participant inclusion. As it was not necessary to discard more than 15% of data for any participant, all participants' data were retained for full data analysis.

Figure 3 displays mean response discrepancies during each trial type. As observed in the original Asch paradigm, baseline performance was excellent; response discrepancies during silent trials did not significantly differ from 0 ($t(20) = 0.40$, $p = .696$), where 0 indicates 100% accuracy. Thus, we can be sure that participants could perform the task with an extremely high degree of accuracy when they were not subject to responses of the confederates. Moreover, performance did not significantly differ from 0 during congruent trials for either private ($t(20) = 1.69$, $p = .106$) or public ($t(20) = 0.76$, $p = .456$) conditions. However, as predicted, response discrepancies were significantly larger, indicating conformity towards the confederates' responses during incongruent trials relative to congruent trials in both the private (private congruent response discrepancy = 0.57, standard error of the mean [SEM] = 0.34, private incongruent = 2.10, SEM = 0.36; $t(20) = 3.10$, $p = .006$, $d = 0.96$) and public conditions (public congruent = -0.26, SEM = 0.34, public incongruent = 4.07, SEM = 0.41; $t(20) = 7.79$, $p < .001$, $d = 2.51$).

Data were entered into a repeated-measures two-way analysis of variance (ANOVA), with condition (private or public) and congruency (congruent or incongruent) as within-subjects factors. This revealed no significant overall difference between performance in private and public trials ($F(1,20) = 3.67$, $p = .070$, $\eta_p^2 = .16$), but a significant main effect of congruency ($F(1,20) = 69.76$, $p < .001$, $\eta_p^2 = .78$) and crucially, a significant

condition*congruency interaction ($F(1,20) = 12.66$, $p = .002$, $\eta_p^2 = .39$), showing congruency effects to be significantly larger during public compared to private trials.

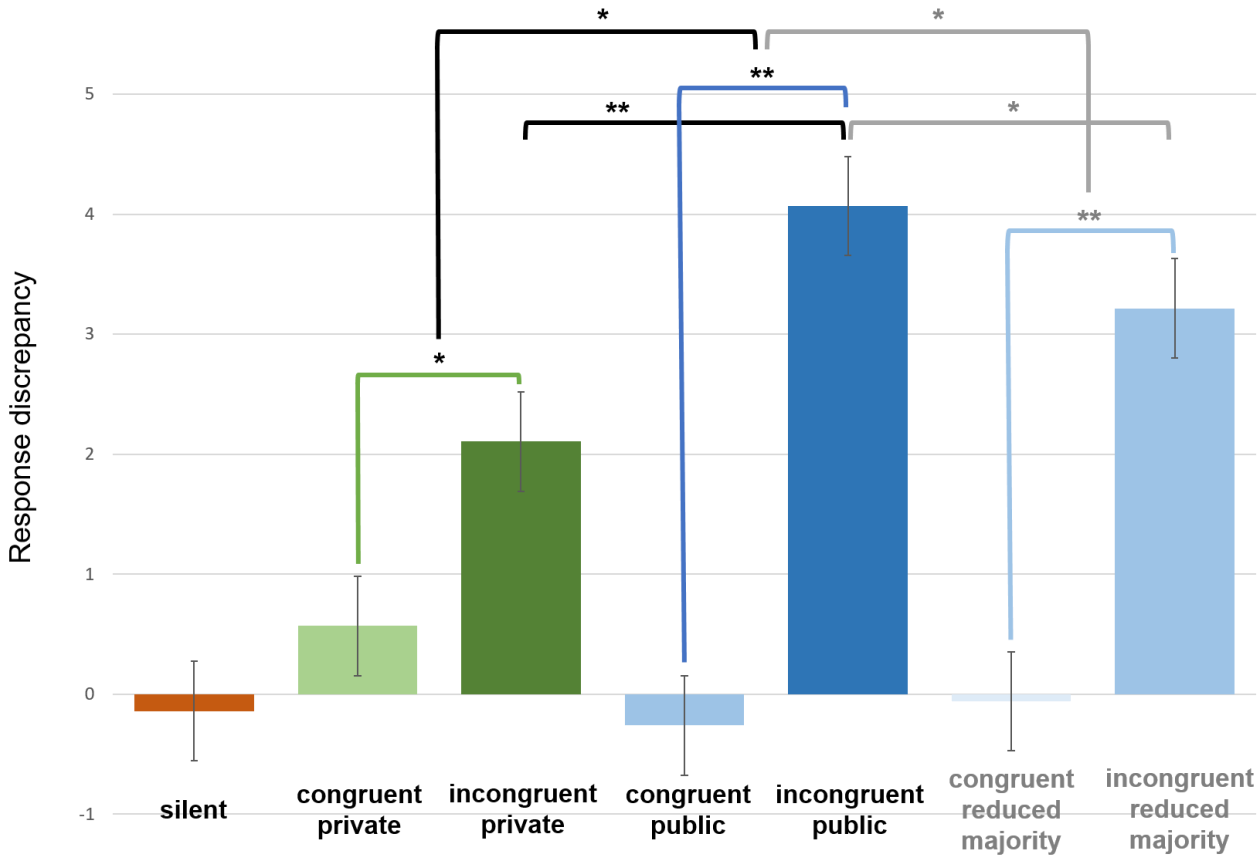

Figure 3. Bar graph representing mean response discrepancies for each trial type. ** indicates significance at $p < .001$ and * indicates significance at $p < .05$. Grey lines show comparisons between standard public trials and reduced majority public trials.

The conformity effect during private trials can be considered to arise only from informational influence and this effect is indeed significantly different from 0 ($t(21) = 3.10$, $p = .006$). However, the additional conformity effect observed on public trials (public minus private effect) can be accounted for by normative influence, and is also significantly different from 0 ($t(21) = 3.56$, $p = .002$). Thus, we see the significant presence of both informational and normative influence. Moreover, it is possible to compare the relative magnitude of

informational influence (indexed by the private conformity effect; mean = 1.53, SEM = 0.50) and normative influence (indexed by public conformity minus private conformity effects; mean = 2.80, SEM = 0.79). Although the normative influence is numerically larger than that of informational influence, the normative influence was not significantly greater in magnitude than informational influence when assessed statistically ($t(20) = 1.06$, $p = .303$, $d = 0.42$). Power analyses conducted with the observed effect size indicate an increased sample size of 40 participants would be sufficient to see a significant difference between the size of the effects (alpha = 0.05, beta = 0.8).

A significant congruency effect was also observed on public trials when responses of only two confederates were available to the participant (i.e. on reduced majority trials; $t(20) = 9.05$, $p < .001$, $d = 2.73$), whereby response discrepancies were larger during incongruent (mean = 3.21, SEM = 0.32) than congruent trials (mean = -0.59, SEM = 0.19). To examine the impact of the size of the majority (number of confederate responses available to the participant) on performance during public trials, we compared congruency effects as well as response discrepancies from incongruent trials between public trials where responses of 2 versus 3 confederates were available to the participant. Response discrepancies during incongruent public trials were significantly larger when the majority consisted of 3 confederates (mean = 4.07, SEM = 0.41) than 2 confederates (mean = 3.21, SEM = 0.32; $t(20) = 2.72$, $p = .013$, $d = 0.51$). Moreover, there was a significantly increased congruency effect under a majority of 3 (mean = 4.33, SEM = 0.56) compared to 2 (mean = 3.27, SEM = 0.36; $t(20) = 2.29$, $p = .033$, $d = 0.49$) confederates.

Finally, the magnitude of informational and normative influence can be compared for trials in which only two confederates' responses were available to the participant. Once again, the normative influence here (mean = 1.74, SEM = 0.63) is numerically, but not statistically, larger than that of informational influence (mean = 1.53, SEM = 0.50; $t(20) =$

0.19, p = .851). The normative influence arising from 3 confederate trials was significantly greater than the normative influence arising from 2 confederate trials ($t(20) = 2.29$, p = .033, d = 0.32).

7.4. Discussion

Compliance and acceptance are types of social influence which can be discriminated on the basis of their effect on the beliefs or attitude of the individual. Compliance occurs when the individual publicly accepts the group's position but privately adheres to their own belief, while acceptance occurs when the individual internalises the belief or attitude of the group such that it becomes their own. Compliance is thought to be the result of normative influence; where the individual complies with group norms due to the social power of the group. Acceptance is thought to be the result of informational influence; where individuals seek information from others in order to determine the true state of the world. The impact of normative and informational influence can be determined through on-line conformity paradigms by comparing responses made by individuals in private and in public. Here we report a novel paradigm, based on that of Asch (1951, 1955), in which both normative and informational influence can be measured on a within-subjects basis. Results indicated significant conformity effects (congruency effects) during both private and public conditions, and crucially, a greater conformity effect during public than private trials, demonstrating the presence of both informational influence in the private condition and the addition of normative influence during public conditions. Therefore, we see demonstrations of both acceptance and compliance within the same task.

The preliminary data reported here suggest that the paradigm is a valid test of social conformity. Based on Asch's (1951, 1955) original paradigm, it compared performance on a test of colour matching in a baseline condition when participants were unaware of the group's

judgements, when they were aware of the responses of two or three confederates but could give their own response in private, and when participants were aware of the responses of two or three confederates and had to give their responses in public. In common with Asch's original findings, participants conformed to the judgements of others in the group; despite the task being sufficiently easy (participants' responses were extremely accurate when not influenced by the group), responses were more inaccurate when the group gave incorrect responses. Also in common with the findings of Asch was the observation that the degree of conformity was greater when participants were exposed to three confederates than two. This was observed even though the 'degree of conformity' in the original Asch paradigm refers to the frequency of conformity, whereas here it refers to the magnitude of the conformity effect on a per trial basis.

The paradigm reported here has the advantage that measures of normative and informational influence are obtained from the same individual at the same point in time, facilitating sensitive comparisons as to their magnitude and enabling future studies using techniques such as fMRI which are reliant upon within-subject comparisons. In addition, the participant is not excluded from the group due to their private responses, avoiding the possibility of anti-conformity to the group judgements which may have influenced prior studies (Abrams et al., 1990). Finally, the use of a continuous response scale allowed extremely subtle effects to be detected.

While the present results are encouraging, and support the validity of the task as a measure of conformity and its ability to measure informational and normative influence, it should be noted the results were obtained with a relatively small sample size and are therefore in need of further replication. In addition, as the current study was primarily aimed at task development and validation, variance due to key individual differences was deliberately reduced. For example, due to the literature on sex differences in social

conformity (Cooper, 1979; Eagly, Wood, & Fishbaugh, 1981; Larsen, Triplett, Brant, & Langenberg, 1979), both concerning the sex of the participant and also whether the group is of the same or opposite sex to that of the participant, the current study included only female participants and a group of female confederates. Moreover, we restricted our sample to young adults (aged 19-30 years) to reduce age-related individual differences.

It can be seen then, that although these results support the use of the task to measure compliance and acceptance, types of social influence thought to arise from normative and informational influence respectively, further work is required to establish the replicability of these findings, and how they are moderated by factors such as age, sex, and culture.

7.5. Conclusions

Here we present preliminary data validating a task designed to identify compliance and acceptance as a result of group influence, and therefore to identify normative and informational influences on decision-making. Although heavily based on the conformity paradigm developed by Asch (1951, 1955), the task has several advantages over previous versions of the Asch paradigm. Firstly, public and private responding is manipulated on a within-subjects basis rather than between-subjects, providing a more sensitive measure of any difference in social influence in the two conditions. This feature also allows participants to respond in the same manner as confederates, reducing the likelihood that they will classify themselves as an out-group member. Secondly, participants were able to respond on a continuous scale – this enabled a graded measure of conformity to be established such that the magnitude of any effect on a single trial could be measured and subtle effects of conformity detected. Results supported previous claims of both normative and informational influences in Asch-type paradigms, with normative influence increasing with increasing majority size in such paradigms.

7.6. Supplementary Data Analysis for Chapter 7

Chapter 7 describes a task designed to investigate higher-level, group social interaction, whilst still eliciting conflict between the self and others. To place this chapter and the task in the context of this thesis, these self-other interference effects on performance during the simple visual perception task can be seen during both private and public social conformity settings. More explicitly, conformity was higher during trials in which the confederates' responses were incongruent compared to congruent with the correct response in both private and public conditions, and the interference effect was greater during the public than private condition. This demonstrates both a form of informational influence which occurs during private situations, as well as an additional normative influence which occurs during public situations. However, in line with the current thesis and findings from Chapters 2 (Sowden, Koehne, Catmur, Dziobek, & Bird, 2016), and 6 whereby autistic trait severity appears to be related to the size of these self-other interference effects (whether in the domain of imitation-inhibition or lie detection), it is of interest to assess the same question with the current task.

Due to time restraints, the shortened 10-item Autism Spectrum Quotient (AQ-10) was given to all participants prior to completion of the task. However, previous research has demonstrated that the AQ-10 is as successful in classifying Autism Spectrum Disorder as the 50-item version of the AQ (Booth et al., 2013). Bivariate correlations reveal a significant positive correlation between AQ-10 scores and the public conformity effect under our one-tailed prediction based on our previous studies (r = .396, p = .038) but not the private conformity effect (r = -.153, p = .254). Moreover, under a one-tailed prediction, there was a significant positive correlation between AQ-10 scores and the difference between public and private conformity effects (r = .377, p = .046). This difference score is reflective of the

additional normative influence over and above that of informational influence. See Figure 4 for scatterplots of these relationships.

Although underpowered to detect anything but large effects, this pilot data does suggest promise for the use of the task in the future investigation of self-other control and its relationship to autistic traits. The data are suggestive of a selective relationship between autistic traits and self-other interference effects specifically in public situations, as well as specific to normative, rather than informational influence. This may suggest different and distinct levels of self-other control, dependent on how an individual (the self) is influenced by their behaviour being publicly available to the group.

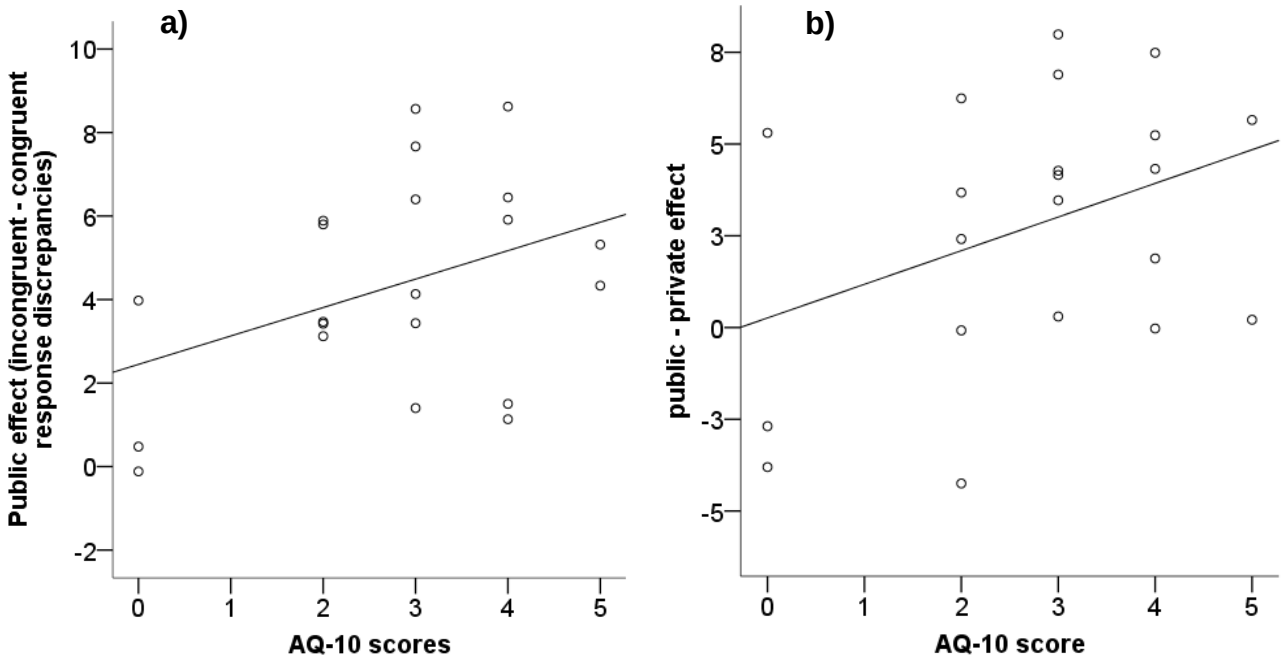

Figure 4. Scatterplots of the relationship between **a)** AQ scores and the public effect and **b)** AQ scores and the public minus private effect.

7.7. References

Abrams, D., Wetherell, M., Cochrane, S., Hogg, M. A., & Turner, J. C. (1990). Knowing what to think by knowing who you are: Self- categorization and the nature of norm formation, conformity and group polarization. *British Journal of Social Psychology, 29*, 97-119.

Allen, V. L. (1965). Situational factors in conformity. *Advances in Experimental Social Psychology*, *2*, 133-175.

Allen, V. L. (1975). Social support for nonconformity. *Advances in Experimental Social Psychology*, *8*, 1-43.

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzdow (Ed.), *Groups, leadership and men: Research in human relations* (pp. 295-303). Lancaster, UK: Carnegie Press.

Asch, S. E. (1955). Opinions and social pressure. *Scientific American*, *193*, 31-35.

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, *70*, 1-70.

Berns, G. S., Chappelow, J., Zink, C. F., Pagnoni, G., Martin-Skurski, M. E., & Richards, J. (2005). Neurobiological correlates of social conformity and independence during mental rotation. *Biological Psychiatry*, *58*, 245-253.

Bond, R., & Smith, P. B. (1996). Culture and conformity: A meta-analysis of studies using Asch's (1952b, 1956) line judgement task. *Psychological Bulletin*, *119*, 111-137.

Booth, T., Murray, A. L., McKenzie, K., Kuenssberg, R., O'Donnell, M., & Burnett, H. (2013). Brief report: An evaluation of the AQ-10 as a brief screening instrument for ASD in adults. *Journal of Autism and Developmental Disorders*, *43*, 2997-3000.

Campbell-Meiklejohn, D. K., Kanai, R., Bahrami, B., Bach, D. R., Dolan, R. J., Roepstorff, A., & Frith, C. D. (2012a). Structure of orbitofrontal cortex predicts social influence. *Current Biology*, *22*, 123-124.

Campbell-Meiklejohn, D. K., Simonsen, A., Jensen, M., Wohlert, V., Gjerløff, T., Scheel-Kruger, J., ... & Roepstorff, A. (2012b). Modulation of social influence by methylphenidate. *Neuropsychopharmacology*, *37*, 1517-1525.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, *55*, 591-621.

Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. In D. T. Gilbert, S. T. Fiske & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed.) (pp. 151–192). Boston, MA: McGraw-Hill.

Cooper, H. M. (1979). Statistically combining independent studies: A meta-analysis of sex differences in conformity research. *Journal of Personality and Social Psychology, 37*, 131-146.

Crutchfield, R. S. (1955). Conformity and character. *American Psychologist*, *10*, 191-198.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgement. *Journal of Abnormal & Social Psychology, 51*, 629-36.

Eagly, A. H., Wood, W., & Fishbaugh, L. (1981). Sex differences in conformity: Surveillance by the group as a determinant of male nonconformity. *Journal of Personality and Social Psychology*, *40*, 384.

Kelman, H. C. (1958). Compliance, identification, and internalization three processes of attitude change. *Journal of Conflict Resolution, 2*, 51-60.

Klucharev, V., Hytönen, K., Rijpkema, M., Smidts, A., & Fernández, G. (2009). Reinforcement learning signal predicts social conformity. *Neuron, 61,* 140-151.

Klucharev, V., Munneke, M. A., Smidts, A., & Fernández, G. (2011). Downregulation of the posterior medial frontal cortex prevents social conformity. *Journal of Neuroscience, 31*, 11934-11940.

Larsen, K. S., Triplett, J. S., Brant, W. D., & Langenberg, D. (1979). Collaborator status, subject characteristics, and conformity in the Asch paradigm. *The Journal of Social Psychology*, *108*, 259-263.

Levy, L. (1960). Studies in conformity behavior: A methodological note. *The Journal of Psychology*, *50*, 39-41.

Nail, P. R., Di Domenico, S. I., & MaCdonald, G. (2013). Proposal of a double diamond model of social response. *Review of General Psychology*, *17*, 1-19.

Sowden, S., Koehne, S., Catmur, C., Dziobek, I., & Bird, G. (2016). Intact automatic imitation and typical spatial compatibility in autism spectrum disorder: Challenging the broken mirror theory. *Autism Research*, *9*, 292-300.

Turner, J. C. (1985). Social categorization and the self-concept: A social cognitive theory of group behaviour. In E. J. Lawler (Ed.), *Advances in group processes: Theory and research* (pp. 77-121), Vol. 2, Greenwich, CT: JAL Press.

Turner, J., T. (1991). *Social influence*. California, CA: Brooks Cole Publishing Co.

Zaki, J., Schirmer, J., & Mitchell, J. P. (2011). Social influence modulates the neural computation of value. *Psychological Science, 22*, 894-900.

8. General Discussion

8.1. Summary of Data Chapters

It is intuitive that during any social interaction, the actions, beliefs, intentions, speech or affective state of the person we are interacting with will have an influence on ourselves. The current thesis aimed to investigate the mechanism by which the impact of both the other on the self, and the self on the other, can be controlled. The thesis focussed on the complex requirement for individuals to represent, and deal with conflict between, both the self and other(s) in various social domains. This thesis therefore examined whether a mechanism coined 'self-other control' may help to explain both typical and atypical social cognitive function. There were three main themes to this research which included the use and development of behavioural indices of self-other control across domains of social cognition, utilising these tasks to assess the integrity of self-other control in relation to disorders of social cognition, and finally to further examine the neural basis of such a mechanism for social cognition.

A brief summary and interpretation with respect to the thesis aims will now be given for each chapter separately, as well as discussion of the strengths and limitations associated with each chapter.

8.1.1. Chapter 2

8.1.1.1. Summary and interpretation. Chapter 2 (Sowden, Koehne, Catmur, Dziobek, & Bird, 2016a) examined imitative behaviour in two samples of individuals with and without a diagnosis of Autism Spectrum Disorder (ASD), in order to establish whether the Broken Mirror Theory of ASD is a plausible explanation for the social impairments experienced by these individuals. As highlighted in this chapter, there are a number of flaws in the neuroscientific investigation of the mirror neuron system (MNS; Hamilton, 2013) and

thus behavioural measures of automatic imitation provide one of the most promising measures of MNS integrity, as automatic imitation is thought to be mediated by the MNS (Catmur, Walsh, & Heyes, 2009; Heiser, Iacoboni, Maeda, Marcus, & Mazziotta, 2003).

Using a well-controlled experimental task, Experiment 1 provided the first large scale demonstration that automatic imitative behaviour, and thus the functioning of the MNS, is intact in individuals with ASD and a matched sample of neurotypical control participants. Moreover, Experiment 2 introduced the key addition of a non-social control variant of the task, whereby a more general cognitive control ability could be measured. Using this isolated, purer measure of imitative control, Experiment 2 demonstrated that imitative behaviour was not only intact in a group of individuals with clinical diagnoses of ASD but in fact appeared to be increased with increasing autism symptom severity. This association was selective to the social, and not the non-social, element of the task and thus to the process involving the control of self and other motor representations.

Therefore, with respect to the main question of the current thesis, this evidence suggests that simple representation of the 'self' and 'other' is indeed intact in individuals with ASD, but that the increased imitation observed here may reflect an impairment in the ability to control or modulate these representations online during the task. Accordingly, these data support the idea that individuals with ASD may experience impairments in 'self-other control', and that these cannot be attributed to a more general executive or attentional control process.

8.1.1.2. Strengths and limitations. Experiment 1 provides the first investigation of automatic imitation in a large and therefore adequately-powered sample. Where previous investigations may have been underpowered, this sample should be sufficiently powered to

detect any differences in imitative behaviour between groups should they exist, and thus is a valuable replication and extension of previous findings.

Moreover, the experimental tasks of imitation used here are good as they measure a form of automatic, involuntary imitation, which controls for other high-level cognitive and motivational processes known to differ between ASD and neurotypical individuals, such as working memory, executive functioning and pragmatic language understanding (Courchesne et al., 1994; Happé & Frith, 1995; Russell, 1997; Williams, Goldstein, Carpenter, & Minshew, 2005). These are cognitive abilities inherent in previously employed voluntary imitation measures (Williams, Whiten, & Singh, 2004) and thus the current study provides a purer measure of imitation. Moreover, both tasks employ a non-social control element to the task, which is an improvement on the original imitation-inhibition task. For example, the task from Experiment 1 included a control task (termed effector compatibility) whereby on half of trials, rather than a finger lift being observed, the finger was simply highlighted by a green mask, drawing attention to the relevant effector but without a finger lift which would elicit a neural motor representation of the action in the participant's own brain.

Furthermore, the task from Experiment 2 took this one step further to create the most recently revised version of the imitation-inhibition task (Sowden & Catmur, 2015), which includes rigorous experimental control over the measure of imitative compatibility but also a non-social, spatial compatibility variant on the task. The main advantage of this task is that both spatial and imitative compatibility are measured during each trial of the experiment with precisely the same experimental stimuli, making them perfectly matched to one another. Each level of imitative compatibility (compatible/incompatible) is measured at each level of spatial compatibility (compatible/incompatible), eliminating the chance of different task demands or stimuli between the two key conditions influencing the results. This also allows conclusions to be made regarding the specificity of the relationship between ASD severity and imitative

control and not to just a general form of executive or attentional control known to be impaired in these individuals (Bird, Catmur, Silani, Frith, & Frith, 2006; Courchesne et al., 1994).

Nevertheless, it could be important to consider whether there is another level on which to distinguish imitative and spatial compatibility in task 2 which is not related to the social versus non-social element of the task. It may be that the selectivity of the mapping between the stimulus and the response distinguishes these two measures in their relationship with ASD severity. For example, at the imitative level, mappings may be considered to be one-to-one, whereby the representation of the other's action (index finger lift) maps onto one's own representation of that action at a high level of specificity, thus promoting a specific index finger lift response in the observer. However, at the spatial level, the observation of a movement on the left side of body space activates a far less selective motor representation for any movement on the left side of the body. It may be that ASD severity is therefore associated with cognitive control performance when there are one-to-one mappings between the stimulus and response, as opposed to the social versus non-social nature of the representations elicited. Further investigations to pick apart imitative and spatial compatibility may help to shed light on this potential alternative explanation.

Additionally, it should be taken into account that although the use of the current tasks allowed for a high degree of control over additional cognitive demands inherent to more high-level tasks, these may be considered to lack ecological validity. The stimuli used include only the presence of hands (in isolation from the rest of the human body) and these are portrayed in two positions only; a static (at rest) position which transitions to an action performance position where a finger is lifted in isolation from the rest of the hand. This is not typical of the way we observe hand actions in everyday social interactions and thus caution should be taken in extending these results to all aspects of everyday imitative behaviour.

However, at this early stage of investigation, it is beneficial to have a high degree of experimental control over extraneous variables, to isolate the measurement of the processes which are and are not impaired in individuals with ASD. It may also be posited that, in a study investigating the control of self and other representations, stimuli presented from a third-person perspective (e.g. rotated by 180°) would be more appropriate than the current hand stimuli which are presented from a first-person perspective. Participants may associate these more with the self and thus a more self-referential process may be elicited here. Third-person perspective hand stimuli may be more successful in eliciting true representations of the 'other' in future investigations.

Finally, this chapter assesses the plausibility of the Broken Mirror Theory of brain function in ASD. However, no measure of neuronal function was included and therefore the integrity of the MNS in these individuals is inferred from a behavioural measure as a well-established product - but an indirect index - of MNS activity. It may be better to combine this task with a neuroscientific method. Conversely, at present, assessments of neuroimaging investigations of MNS function in ASD reveal huge inconsistencies (Hamilton, 2013) and therefore cast doubt on the validity of these methods. Instead, causal evidence from brain stimulation maybe be more fruitful in combination with these behavioural indices of MNS function.

8.1.2. Chapter 3

8.1.2.1. Summary and interpretation. Chapter 3 (Sowden, Brewer, Catmur, & Bird, 2016b) of the current thesis investigated the relationship between alexithymic traits and the integrity of self-other control via the same revised imitation task used in Experiment 2 of Chapter 2 (Sowden & Catmur, 2015). Recent evidence predicts an association here as alexithymia is purported to be characterised by impairments in one's ability to perceive the

internal state of one's own body (Brewer, Cook, & Bird, 2016a, 2016b; Murphy, Catmur, & Bird, in press), which is suggested to be crucial for awareness of one's own body representations and its distinction from the body representations of others (Seth, 2013; Quattrocki & Friston, 2014). Thus, in line with recent evidence of the relationship between imitation-inhibition (as an index of self-other control) and interoceptive ability (Ainley, Brass, & Tsakiris, 2014), as well as evidence for the relationship between alexithymia and interoceptive ability (Brewer et al., 2016b; Murphy et al., in press; Shah, Hall, Catmur, & Bird, 2016), a relationship was predicted between imitation-inhibition and alexithymia in the current study.

The results from this chapter support this prediction, whereby increasing alexithymia was associated with an improved ability to inhibit imitation, with precisely the same specificity as imitation-inhibition has previously been found to associate with interoceptive ability (Ainley et al., 2014). For example, the relationship was not only specific to imitative, and not spatial, compatibility (as a non-social, attentional control variant), but was also specific to imitatively incompatible trial performance. This suggests the relationship is driven by the ability to distinguish and control representations of one's own motor representation from that of the other when the self and other are in conflict, rather than being driven by the basic tendency to imitate. Moreover, alexithymia showed a specific predictive relationship here, which was independent of any relationship with autistic traits when measured by Autism Spectrum Quotient (AQ) scores.

Therefore, a few conclusions can be drawn from the present study. Firstly, in conjunction with the data by Ainley and colleagues, the data support the suggestion that alexithymia may not only be characterised by impairments in interpreting emotions of the self and others but rather, by a general deficit in interoception; the ability to perceive and interpret the internal representational state of both the self and others. This may help to explain the

broad co-occurrence of alexithymia with a number of disorders of social cognition such as autism, schizophrenia and eating disorders (Brewer et al., 2016a; Murphy, Brewer, Catmur, & Bird, 2017). Additionally, these findings are significant for our understanding of the specificity of intercoceptive and behavioural profiles of individuals with high levels of alexithymia, and necessitate revision of current theoretical models of the relationship between interoception, self-other control and social cognitive ability.

Finally, with respect to the main theme of the current thesis, these data extend our understanding of the role of self-other control across disorders of social cognition. They begin to reveal the important role for self-other control in the social impairments associated with alexithymia, and its independence from autistic traits.

8.1.2.2. Strengths and limitations. The strengths and limitations of the imitation task employed in this chapter are as outlined with respect to the task in Experiment 2 of Chapter 2. Specific to Chapter 3, however, the sample is a strength. A population sample was employed, where participants from the full range of alexithymia severities were investigated, providing a comprehensive assessment of the relationship between alexithymia and self-other control processing. Moreover, that an individual differences approach to assessing cognitive abilities as utilised here, is in line with current theories regarding the association between continuous measures of psychological or neuropsychiatric traits and cognitive abilities (Friedman & Miyake, 2017; Graham & Lachman, 2014; Kane et al., 2016).

Nevertheless, it should be taken into account that, although conclusions have been made regarding the relationship between three variables; interoception, alexithymia, and self-other control, only the latter two are measured in the current study. The relationship between the three variables is only indirectly inferred due to a combination of previous strong associations found between; alexithymia and interoceptive ability (Brewer et al., 2016b;

Murphy et al., in press; Shah et al., 2016); interoception and self-other control (Ainley et al., 2014); and in the current study between alexithymia and self-other control (Chapter 3; Sowden et al., 2016b). Thus, it will be important in future research to combine a measure of all three variables within the same study in order to gain a more direct measure of the relationships discussed here.

8.1.3. Chapter 5

8.1.3.1. Summary and interpretation. Chapter 5 (Sowden, Wright, Banissy, Catmur, & Bird, 2015), Experiment 1 details the development of a novel task requiring online self-other control in the context of lie detection, whereby representations refer to the opinions of the self and the other. This video-mediated lie detection task was successful in eliciting self-other interference effects equivalent to those observed in the imitation-inhibition task, with participants being significantly better at detecting truths and lies during trials in which the opinions of the self and the other were consistent, compared to inconsistent with one another. Experiment 2 then utilised this new index of self-other control to further investigate the involvement of the right temporoparietal junction (TPJ) in this process. As predicted, excitatory transcranial direct current stimulation (tDCS) to the right TPJ reduced this interference effect, improving lie detection performance in situations where the self and other opinions were inconsistent. Taken together, this chapter successfully establishes another domain of social cognition during which self-other control is required; where one must inhibit the self opinion and enhance that of the other person in order to perform the task successfully. By boosting the ability to control competing representations of the self and other – in this case inhibiting one's own opinion in order to more accurately represent that of the other – TPJ stimulation can improve lie detection in situations of inconsistency between the self and other.

8.1.3.2. Strengths and limitations. The task employed in both Experiment 1 and 2 of Chapter 5 is well matched to other tasks of self-other control as it elicits equivalent self-other interference effects – whereby performance is diminished under situations of self-other conflict compared to self-other non-conflict – but in the domain of lie detection where representations of self and other *opinions* are consistent or inconsistent. The task is, however more ecologically valid and is more typical of everyday social interaction than the imitation-inhibition task. As is required in this task, in everyday life we are often exposed to individuals expressing ideas or opinions on a topic and are required to make online decisions about the credibility of the person and the opinion expressed. Missing from this task, however, and present in the imitation-inhibition task, is a separate measure of a non-social control variant on the task. This would allow the investigation of the selective nature of the TPJ's involvement here, specific to '*social'* self-other control and not to a more domain general mechanism of control. However, the specificity of self-other control and indeed the role of the TPJ to social cognition has now been well established via a number of other brain stimulation studies (Hogeveen et al., 2015; Santiesteban, Banissy, Catmur, & Bird, 2012a, 2015; Sowden & Catmur, 2015). Thus, as a task well matched in terms of the requirement for online self-other representation and control, it is likely that the effect is also socially specific here.

In relation to Experiment 2 of Chapter 5 and the use of tDCS to the TPJ, based on the tDCS current best practice guidelines, the use of an *active* control site (namely the mid-occipital control region) is superior to the use of only a sham control condition. The use of an active control site allows greater specificity in assessing the involvement of a particular brain region in the behaviour of interest (Parkin, Ekhtiari, & Walsh, 2015). This form of control is also better matched to the experimental stimulation condition in terms of the physical sensation over the scalp (Russo, Wallace, Fitzgerald, & Cooper, 2013).

It is worth considering the reliability of tDCS here, as well as some reports of lack of replication of cognitive and neurobiological effects of stimulation (Horvath, Forte, & Carter, 2015a, 2015b). Despite arguments that such meta-analyses are flawed in their approaches (Price & Hamilton, 2015), as well as good replicability of and promise for the effects of tDCS in enhancing self-other control (Hogeveen et al., 2015; Santiesteban et al., 2012a, 2015; Sellaro, Nitsche, & Colzato, 2017), some caution should be taken in interpreting the results from Experiment 2 until these have been replicated in an independent sample. A further negative feature relating to the use of tDCS over other methods of neurostimulation such as transcranial magnetic stimulation (TMS) is the reduced focality of stimulation. For example, tDCS electrodes are often 20 or 35 cm^2 in surface area over the scalp, thus stimulating a larger region of the cortex compared to TMS which, when delivered using a figure-of-eight coil, is very focal in its stimulation. Moreover, the spatial resolution of tDCS is often compromised in such studies as stimulation points are based on the electroencephalography 10/20 system and thus are not individually determined based on the participants' own brains. However, previous studies (e.g. Hogeveen et al., 2015; Santiesteban et al., 2012a, 2015) using this same stimulation protocol found specific modulatory effects on social domains requiring self-other control (and not on domain general tasks of control). Thus, it is unlikely that focality or the spatial resolution are limiting factors in the present study. Nevertheless, the use of a more focal method of brain stimulation, such as TMS along with participant-specific neuronavigation (as in Sowden & Catmur, 2015), or even the use of the newly developed high definition tDCS (Edwards et al., 2013), would provide successful avenues for future research to enhance the efficacy of stimulation and the replicability of the results concerning the involvement of the TPJ in self-other control.

On a further critical note, the current study only investigated stimulation over the right TPJ, whereas a number of previous functional magnetic resonance imaging (fMRI) and lesion

studies have also shown the potential importance of the *left* TPJ in social functions (see Saxe, 2009), and even more specifically in self-other control (Santiesteban et al., 2015). Thus, the use of bilateral stimulation may be interesting and would also reduce any potential right lateralised changes in overall neural connectivity. It is also worth noting that despite the present study focusing specifically on the involvement of the TPJ, another brain region argued to be particularly important in the process by which we control and perform social actions is the medial prefrontal cortex (mPFC), established via a number of fMRI studies (Brass, Bekkering, & Prinz, 2001; Brass, Derrfuss, Matthes-von Cramon, & von Cramon, 2003; Brass, Derrfuss, & von Cramon, 2005; Brass, Ruby, & Spengler, 2009; Spengler, von Cramon, & Brass, 2009a, 2009b, 2010). Therefore, it seems intuitive that these areas form a network to fulfil this important function and as a result it is particularly important not to investigate these in isolation, but rather to observe connectivity and information flow between these two regions. A combination of brain stimulation and electroencephalography or fMRI seem appropriate techniques for such future investigations.

8.1.4. Chapter 6

8.1.4.1. Summary and interpretation. Chapter 6 extended the investigation of the integrity of self-other control in ASD to the use of the newly developed lie detection task outlined in Chapter 5. The self-other opinion consistency (or interference) effect found in Chapter 5 was not found in the control group in Chapter 6, and was trending towards significance in the individuals with clinical diagnoses of ASD. However, interestingly, the size of the consistency effect for each individual showed a significant positive association with autistic traits, as measured by the AQ. The greater the consistency effect (i.e. the poorer one's self-other control), the higher the autism trait severity. This was apparent across the whole spectrum of autism trait severity, in both individuals with an ASD diagnosis and those at the low end of the spectrum.

Thus, despite the lack of replication of the basic consistency effect, which will be further discussed in the limitations section, these data provide further evidence for the relationship between performance in another domain of self-other control and autistic traits, further supporting future research into self-other control as a candidate mechanism of atypical as well as typical social cognition.

8.1.4.2. Strengths and limitations. The strengths and limitations of the video-mediated lie detection task employed in this chapter are the same as those outlined for Chapter 5.

The sample size of the ASD group in the current experiment is a limitation and was restricted to the number of individuals with ASD available to be recruited and take part during the data collection period for this study. A trend towards a significant consistency effect with a medium effect size was observed in this group, suggesting that the sample was simply underpowered to detect the effect comparable to that found with the larger sample in Experiment 1 of Chapter 5. Power calculations reveal a sample size of 40 would be sufficient to reveal a significant consistency effect if it were to exist in this group. This does not, however, explain the non-replication of the basic lie detection consistency effect in the control group. Nevertheless, reassessment of the data from Chapter 5, Experiment 1 (Sowden et al., 2015) reveal consistency effects were largely driven by the females in the sample. In combination with our findings here of the relationship between increased autistic traits and increased consistency effects, it appears that consistency effects may be boosted by the presence of ASD as well as status as a female. Thus, a speculative explanation for the lack of a consistency effect found in the control sample may be that the current control sample was made up largely of males, whilst also possessing many individuals at the very low end of the autistic trait spectrum, both of which may have led to decreased or even reversed consistency effects to those observed in Chapter 5, Experiment 1. Unfortunately, AQ scores were not

available in Chapter 5 and thus a direct comparison cannot be made to the present study to ratify this explanation. Therefore, there is need for a large-scale replication of the basic consistency effect prior to future use of the task.

8.1.5. Chapter 7

8.1.5.1. Summary and interpretation. Chapter 7 presents preliminary data from the design and validation stage of a new group social interaction task aimed at separating out two mechanisms of social influence; informational and normative influence (resulting in acceptance and compliance, respectively). The current task was capable of eliciting conformity (or interference) effects, whereby conformity was higher during trials in which the confederates' responses were incongruent compared to congruent with the correct response, in both private and public conditions. This conformity effect was observed to be larger in the public than private condition. This finding suggests the presence of informational influence, whereby an individual accepts a degree of truth in the responses of the others and as a result shifts their opinion towards that of others. However, this also suggests a significant addition of a normative influence effect during public responding trials, whereby public responding leads an individual to comply with the responses of the others in order to fit in, without necessarily holding that response as their true belief.

In the context of the current thesis, we can see two levels of social influence which require an individual to represent both the self and the other(s) and then to control these competing representations in order to respond successfully. Self-other interference effects occur because of the conflict between the self and others, leading to poorer performance when a group of confederates' responses are incongruent with the correct response. In other words, this presents another social domain during which one must inhibit the representation one holds for the others' beliefs and must enhance that of their own in order to perform well

at the task. Moreover, it appears, as was observed in Chapters 2 and 6, that such indices of self-other control may relate to individual differences in autistic trait severity, providing further evidence for self-other control as a candidate mechanism for the impairments across social domains in ASD.

Finally, these data begin to introduce the idea that self-other interference effects, and therefore self-other control, may play a role at many levels of the same social domain. Despite there being higher level distinctions between the forms of social influence occurring, whether it be informational or normative influence (acceptance or compliance), it appears that self-other control may be a low-level mechanism at play across both of these forms of social influence.

8.1.5.2. Strengths and limitations. The current experiment is the first to extend the measurement of self-other performance interference effects to a novel high-level group social interaction task. Moreover, it was demonstrated that self-other interference effects can be investigated and observed at two levels of one social domain; performing a task under influence of a group of others in both public and private situations (when one's response is either accessible or inaccessible to the rest of the group).

Furthermore, the present task represents the first demonstration of conformity under public and private conditions within the same task and thus both were assessed within-subjects with identical task demands and no impact of sampling differences between the two. Therefore, a direct comparison of these effects is afforded and allows conclusions to be made regarding the additional impact of public responding compared to responding in private.

Nevertheless, it must still be appreciated that the current study is simply a small-scale pilot of the task. This requires replication in a larger sample as well as consideration for how the findings are influenced by individual differences known to impact on social conformity

(Bond & Smith, 1996; Crutchfield, 1955) such as sex, age, ethnicity and various personality traits of both the self and others. Here, however, all of these variables were kept as constant as possible to provide a more homogenous group of individuals more likely to behave in a consistent way. For example, only females were considered, as well as only students and staff from King's College London, all aged between 19 and 30 years. Moreover, the task may be considered to lack ecological validity as it was a very simple, low-level visual perception task unlike any task which would be completed in a group social interaction in everyday life. However, this feature of the design was adopted in order to control for extraneous variables inherent in a more complex social interaction task.

8.2. General Summary and Interpretation

In summary, this thesis presents a series of experiments aiming to investigate whether self-other control can be considered a candidate neurocognitive mechanism to explain both typical and atypical social cognitive function. The first theme of the current thesis led to the investigation of behavioural indices of self-other control across domains of social cognition and this body of research successfully extended well-established, low-level tasks such as the imitation-inhibition task to higher level tasks more reflective of complex everyday social and group interaction. Two novel tasks were designed (Chapters 5 and 7), which had the same requirement for online representation and control of the self and other as observed in traditional tasks which index self-other control. The first of these was a lie detection task which involved online representation and control of the opinions of the self and other, and the second task involved representing the responses of the self and a group of others during a simple visual perception task. These novel tasks follow the robust method of inducing self-other performance interference effects, whereby performance is compared between situations when self and other(s') representations were congruent (also termed 'compatible' or 'consistent') and incongruent (also termed 'incompatible' or 'inconsistent') with one another.

As with the very first demonstration of this via the task of automatic imitation (Stürmer, Aschersleben, & Prinz, 2000), interference effects are observed when the representation an individual creates of the action/opinion/visual percept of another individual is conflicting with that of one's own; causing a performance detriment relative to situations when the representation of the self and other(s) are not in conflict.

Moreover, the second theme concerns whether self-other control may be considered a candidate mechanism to explain the broad range of social impairments observed in disorders of social cognition, such as ASD (Sowden & Shah, 2014). Chapters 2, 3, 6 and 7 in the current thesis behaviourally assessed the integrity of self-other control in relation to atypical social cognition, either using case/control designs or by examining traits across the general population. Most focus (Chapters 2, 6 and 7) was on autistic traits and ASD as the prototypical disorder of social cognition, but alexithymia, as another condition argued to be associated with atypicalities in self-other processing was also assessed (Chapter 3). Although, no differences in self-other control were observed between ASD and Control groups in Chapters 2 and 6, this highlights the importance of considering disorders as a spectrum of trait severities. Accordingly, findings from this thesis reveal significant relationships between tasks indexing self-other control and autistic and alexithymic traits. Increased autistic traits were associated with increased self-other interference effects, and thus impaired self-other control across three different tasks, whereas increased alexithymic traits were associated with decreased imitation, and thus superior self-other control. This association with alexithymia was also observed to be independent from any relationship between autistic traits and imitation-inhibition.

Our understanding of the role of self-other control in atypical social functioning is broadened from the data in the current thesis. Despite strong associations between alexithymia and autistic traits as well as between these traits and self-other control, these two

clinical traits appear to be independently and differentially associated with self-other control processes. Nevertheless, Chapters 2 and 3, in their use of the imitation-inhibition task as the index of self-other control, do demonstrate that both clinical traits are selectively associated with the imitative and not the non-social spatial control element of the task. Thus, self-other control may indeed be considered a candidate mechanism contributing to social functioning both across *domains* as well as *disorders* of social cognition (Sowden & Shah, 2014), but it is important to also consider its independent or dissociable role in different disorders.

Additionally, the current thesis further explored the neural basis of self-other control. Specifically, the role of the right TPJ in this mechanism was investigated using tDCS and utilising the novel, higher-level task of self-other control in the context of lie detection (Chapter 5). The findings from this chapter indeed support the role of self-other control and the right TPJ across low and higher-level domains of social cognition.

Finally, the use of the revised imitation-inhibition task (Sowden & Catmur, 2015) in Chapters 2 and 3 allow some conclusions to be made regarding the specificity of self-other control to social cognition. For instance, the task dissociates two control processes within each trial of the experiment; one concerning imitative control (the ability to inhibit the tendency to imitate an action performed by the other) and the other process concerning a non-social, attentional or executive control process (the ability to inhibit the tendency to perform a response which is spatially compatible with an observed action). These two processes have previously been shown to be dissociable in their recruitment of the TPJ (Sowden & Catmur, 2015) and Chapters 2 and 3 also show the selective relationship between the imitative control (but not the spatial control) element of the task and autistic and alexithymic traits. Thus, in conjunction, these data further support the idea that the TPJ and the mechanism of self-other control has a domain specific role in social cognition as opposed to acting as a general inhibitory control mechanism. It is particularly important to consider the selectivity of this

control mechanism to social impairments in ASD as these individuals are consistently reported to show general executive or cognitive control impairments (Courchesne et al., 1994; Happé, Booth, Charlton, & Hughes, 2006; Hill, 2004; Ozonoff, 1997).

8.3. General Limitations

8.3.1. Developing new tasks of self-other control. There are some general limitations in the current body of research which are worth outlining here. The aim was to develop new tasks to index self-other control across different social domains but with the same basic conflict between the self and other, and thus the requirement for online control to be exerted over these representations. However, despite self-other interference effects being observed in each of the domains considered, there are important differences between the tasks which are not controlled for. For example, the imitation-inhibition task requires a speeded response, with reaction time constituting the dependent variable, whereas the lie detection and social conformity tasks do not require speeded responses and performance accuracy makes up the dependent variable of interest. Moreover, the imitation-inhibition task measures a form of involuntary or automatic behaviour, whereas other measures of self-other control including the Director Task (perspective-taking measure), lie detection and social conformity tasks measure a response which may be deliberated over and is a voluntary, planned response. It may be that there is an important distinction between the role of self-other control and the TPJ in involuntary, speeded measures of social cognition versus voluntary measures which is not considered in the current work. These differences between the tasks may explain why tDCS to the TPJ in both the lie detection task (in Chapter 5 of this thesis), and the perspective-taking task (Hogeveen et al., 2015; Santiesteban et al., 2012a, 2015) shows a selective modulation of self-other conflict trials only, whereas the same protocol of stimulation has been shown to modulate both self-other conflict and non-conflict trials of the imitation-inhibition task (Santiesteban et al., 2012a, 2015). All tasks, however, do measure

the relatively unconscious impact of conflict between the representations of the self and other(s) on one's own response and thus the tasks are perhaps sufficiently equivalent at this level to be compared.

Moreover, an important point to consider when developing new tasks of self-other control, is the previous failure to find the modulatory effect of tDCS to the TPJ in two independent theory of mind tasks, as well as a lack of association between other self-other control tasks and theory of mind (Santiesteban et al., 2012a, 2012b 2015). Theory of mind has been reliably found to recruit the right TPJ (Castelli et al., 2000; Saxe, 2009, 2010; Saxe & Kanwisher, 2003; Saxe & Wexler, 2005; Scholz, Triantafyllou, Whitfield-Gabrieli, Brown, & Saxe, 2009), and does intuitively require the key process of self and other representation of mental states, as well as control between these in order for successful theory of mind (Spengler et al., 2009a, 2010). However, experimental tasks used to index theory of mind in this literature include the Strange Stories task (Happé, 1994) and the Movie for the Assessment of Social Cognition (Dziobek et al., 2006) task, both of which do not require online representation and control of the self and other during the response period. For example, both tasks involve listening to or reading about a social interaction between a set of characters after which they are retrospectively asked to reflect on the social situation and answer a set of multiple choice comprehension questions about the behaviour of the characters. Thus, no response is required at the initial stage during which conflict between the mental states of the self and the other(s) occurs, and it may be this initial *online* conflict between self and other during the response period which is crucial for the involvement of the TPJ and indeed this mechanism of self-other control. Future studies may look to design a theory of mind task which elicits equivalent requirements for representation and control of competing self and other representations during the critical response period in order to truly test the involvement of self-other control in theory of mind.

8.3.2. Assessing self-other control in disorders of social cognition. Additionally, there are general limitations when considering investigation of disorders of social cognition. ASD is a particularly heterogeneous disorder making diagnosis difficult. There are also problems with tools used to assess ASD severity in research studies. For example, often measures which provide a severity score are utilised, such as the AQ and the Autism Diagnostic Observation Schedule (ADOS) as is the case in the current thesis. However, in Chapters 2 and 6 as well as in recent assessments of the relationship between AQ / AQ-10 and ADOS severity scores (Ashwood et al., 2016; Morrier et al., 2017) these have been found not to correlate with one another (*p*s > .2). This may be in part due to the AQ being a self-report scale, whereas the ADOS is a diagnostic tool completed by a clinician, but this presents a problem as to which measures should be deemed most valid for use in research. Consequently, the relationship between autistic trait severity and tasks of self-other control in the current thesis are not entirely consistent across AQ and ADOS measures. In Chapter 2, Experiment 2, a correlation was observed between self-other control in the imitation-inhibition task and ADOS severity scores but not AQ severity scores; whereas Chapters 6 and 7 present correlations between self-other control using the lie detection and social conformity paradigms with AQ and AQ-10 scores respectively. No correlation was observed in Chapter 6 with ADOS severity scores.

These results appear contradictory if the tasks used here do indeed index the same underlying mechanism. However, perhaps different elements of an individual's profile of autistic traits are identified and reported by the self, compared to by a clinician, explaining why these may not be associated with one another and may be differentially associated to our self-other control measures. For example, attributes reported by the individual (via the self-report AQ) may be less likely to correlate with a performance measure of automatic or involuntary behaviour (such as the imitation-inhibition task), whereas clinician-rated

attributes of an individual's profile (via the ADOS) are perhaps more likely to relate to these aspects of automatic behaviour which are not apparent to the individual but are visible to a clinician. Likewise, the lie detection and social conformity tasks as previously mentioned, involve more conscious deliberation and knowledge of one's own responses and thus may be more likely to relate to self-reported than clinician-reported attributes of an individual's autistic profile.

The co-occurrence between disorders can also present a problem in research due to the difficulty in separating out the contribution of one disorder from another on the behavioural trait of interest. For example, depression and anxiety regularly co-occur with ASD, and a body of work has now demonstrated that the emotional symptoms thought to be a core feature of ASD, schizophrenia and eating disorders are in fact accounted for by alexithymia as a separate but co-occurring condition (Brewer, Cook, & Bird, 2014; Brewer, Cook, Cardi, Treasure, & Bird, 2015; Cook, Brewer, Shah, & Bird, 2013). Thus, it is important to consider and measure co-occurring disorders or conditions when investigating any neurodevelopmental or neuropsychiatric condition. Chapters 3 and 6 of the current thesis indeed included regression analyses which examined the unique contribution of alexithymic and autistic traits to performance in two different tasks of self-other control. Future research in this area would benefit from a combination of population representative samples but also targeted, highly controlled studies to try to investigate disorder-specific impairments and those which may be present due to other co-occurring conditions. Additionally, studies should include measures of traits commonly associated with neuropsychological conditions such as depression, anxiety and self-esteem.

Finally, it is particularly important to recognise that the mechanism of self-other control here does not replace other theories of the social deficits observed in disorders such as ASD, but rather may compliment these as an underlying and lower-level mechanism to

explain consistency between performance in different social domains and between social disorders characterised by similar social impairments. For example, one theory which appears now to lack plausibility (Hamilton, 2013) in ASD is the Broken Mirror Theory (Iacoboni & Dapretto, 2006; Ramachandran & Oberman, 2006). Other theories regarding the social impairments in ASD, which were beyond the scope of the current thesis to investigate, also include the Social Motivation Theory (Chevallier, Kohls, Troiani, Brodkin, & Schultz, 2012), the Theory of Mind Hypothesis (Baron-Cohen, Leslie, & Frith, 1985) and the Social Attention or Social Orienting Hypothesis (Leekam, 2016; Senju & Johnson, 2009). These theories may go a little way to explaining some of the specific problems individuals with ASD have in understanding and performing well in social situations. However, the mechanism of self-other control may have the capacity to explain why some basic automatic social behaviours are intact in these individuals whilst they lack control or modulation of their social behaviour across broad social domains. In line with the conceptual shift towards viewing ASD as a spectrum disorder, this mechanism allows the consideration of individual differences in social cognitive functioning across the spectrum of autistic traits which are difficult to account for under other theories which take a categorical view on neuropsychiatric disorders.

Nevertheless, it is important to appreciate that self-other control is just one mechanism contributing to the extremely complex pool of processes underlying social functioning. Accordingly, along with the multi-factorial nature of clinical disorders, we cannot presume a single cause or mechanism to wholly explain atypical social cognitive development, and one must also consider the direction of causality here; whether deficient self-other control may be a cause or a consequence of neurodevelopmental disorders.

Finally, an important consideration of the current body of research is that of sample sizes. For example, the sample sizes of individuals with ASD used in the current thesis

(Chapter 2, Experiment 2 and Chapter 6) are restricted to the number of individuals who had consented in previous experiments to be re-contacted and were then available to take part during the testing period. However, a notable problem with this form of recruitment is that sample size cannot always be determined a priori to enable enough power to detect a prespecified effect size (Farrokhyar, Reddy, Poolman, & Bhandari, 2013). A small sample size may simply be underpowered to detect a true effect (type II effect), but also, via the potential introduction of systematic bias or error may lead to a false positive effect (type I error). Therefore, determining the appropriate sample size a priori and recruiting and testing a specified number of individuals would help to mitigate these concerns (Faul, Erdfelder, Lang, & Buchner, 2007).

8.4. Outstanding Questions

8.3.1. Applicability to other disorders of social cognition: Schizophrenia. We have seen in the current thesis that the mechanism of self-other control may underlie many facets of social functioning, but also that it may be considered a mechanism which is impaired in individuals with disorders of social cognition, such as ASD and alexithymia. However, it is important to consider how this may extend to other neuropsychiatric disorders. One of particular interest is that of schizophrenia. A short outline of the potential for self-other control in the context of this disorder will now be considered.

As with ASD, schizophrenia has been associated with various impairments in social functioning usually present prior to and persisting after the onset of symptoms (Tandon, Keshavan, & Nasrallah, 2008). Behaviourally, individuals with schizophrenia are often described as having diminished self-reflection (Ferri et al., 2012; Roe & Davidson, 2005; van der Meer, Costafreda, Aleman, & David, 2010), abnormalities in attributing agency to the self and others (Graham-Schmidt, Martin-Iverson, & Waters, in press; Renes, Vermeulen, Kahn,

Aarts, & van Haren, 2013) and more general impairments in self-other integration and distinction, which have been argued to account for the broad social cognitive deficits associated with schizophrenia (Langdon et al., 1997; van der Weiden, Prikken, & van Haren, 2015). Moreover, as with ASD, key aspects of the symptom profile in schizophrenia may be explained by a deficit in self-other control. For instance, identity and reality disturbances such as hallucinations and thought insertion exemplify a misattribution of self-generated, internal representations to others or the external world, highlighting a difficulty in managing representations of self and others (Allen, Freeman, Johns, & McGuire, 2006; Allen et al., 2004; Jeannerod, 2009).

It is now well established that fronto-temporal functional connectivity is reduced in schizophrenia and this has been linked to diminished top-down modulatory control over social behaviour (Allen, Larøi, McGuire, & Aleman, 2008; Cook, Barbalat, & Blakemore, 2012). Abnormal structure and function of the TPJ as a key node within this network has been reported in individuals at risk (Brüne et al., 2011), as well as those suffering from schizophrenia (Benedetti et al., 2009; Brüne et al., 2008; Das, Lagopoulos, Coulston, Henderson, & Malhi, 2012; de Achával et al., 2012; Koeda, Takahashi, Matsuura, Asai, & Okubo, 2013; Lee, Quintana, Nori, & Green, 2011), relative to healthy controls. Diminished activation of this region has been associated with impaired social cognitive performance, particularly in domains such as theory of mind and emotion processing. Moreover, abnormal structure of the inferior parietal lobule (Torrey, 2007) and function of the superior temporal gyrus (Allen et al., 2004, 2006; Koeda et al., 2013; Mechelli et al., 2007) have been associated with social functioning deficits, particularly in judging agency and symptoms such as visual and auditory hallucinations. Both the posterior end of the superior temporal gyrus and the most inferior aspect of the inferior parietal lobule border the TPJ.

Taken together, evidence to date is suggestive of the TPJ's role in a common neurocognitive mechanism which may be deficient across disorders. Here we have considered how impairments associated with schizophrenia as well as ASD and alexithymia may fit into a framework of atypical self-other control, paving the way for future research in this area. Such a mechanism may also aid in our understanding and treatment of the cognitive symptoms experienced by individuals with disorders of social cognition. They also raise the possibility that cross-disorder cognitive interventions could be designed incorporating training in self-other control.

8.3.3. The future for the neural basis of self-other control. There is now a large body of evidence to support the involvement of the TPJ in the mechanism of self-other control and this indeed appears to be specific to social cognition and not to other more general inhibitory processes (Hogeveen et al., 2015; Santiesteban et al., 2012a, 2015; Sowden & Catmur, 2015; Sowden et al., 2015). However, it is important to consider the TPJ's role in different functional networks in the brain and indeed the different input and output organisation of this region in the context of each of these networks and how they relate to social and non-social function (Krall et al., 2015; Mars et al., 2012). Moreover, another region suggested to be involved in this process is the mPFC (Brass et al., 2005, 2009; Costa, Torriero, Oliveri, & Caltagirone, 2008; Spengler et al., 2009a). The TPJ is purported to be particularly important in determining agency (whether the action is performed by the self or the other), whereas the mPFC primarily manages and enforces our own motor representations and actions as a result (Brass et al., 2005, 2009). Therefore, it seems intuitive that these areas form a functional network to fulfil this important function, but further research is required to confirm the individual and combined roles of these regions in facilitating self-other control. No individual brain region functions in isolation and consequently the use of a combination of methods (e.g. neuroimaging and neurostimulation) will help to establish the network

within which the TPJ functions, as well as its connectivity with the mPFC. Moreover, this will allow the investigation of the consequence of exciting or inhibiting one region on the integrity of the whole network as well as behavioural measures of self-other control.

Additionally, this body of work concerning the neural basis of self-other control via neurostimulation to the TPJ is yet to be investigated in clinical populations which is a fruitful avenue for future research. Although a short session of neurostimulation has only short-lasting aftereffects (up to 90 minutes), there is evidence to suggest that repeated sessions at intervals of around 24 hours (Reis et al., 2009) and a combination of tDCS and cognitive behavioural training may have much longer lasting effects of 2 months or more (Looi & Cohen Kadosh, 2016). Such an intervention may pave the way for future understanding and application of the mechanism of self-other control to atypical social functioning, and much interest and recent success has been seen in the utility of tDCS to treat depression (Nitsche, Boggio, Fregni, & Pascual-Leone, 2009; Shiozawa et al., 2014) and some other clinical conditions such as stroke, Parkinson's disease and pain-related conditions (Borckardt et al., 2011, 2012; Demirtas-Tatlidede, Vahabzadeh-Hagh, & Pascual-Leone, 2013; Fregni et al., 2006; Lindenberg, Renga, Zhu, Nair, & Schlaug, 2010). This work is yet to be extended to the enhancement of social cognitive function.

Finally, it will also be important to consider other suggested roles for TPJ in social cognition. For example, another competing idea is that the mPFC and TPJ, rather than facilitating the control of competing representations of self and other, may in fact help to differentiate task-relevant from task-irrelevant representations (Cook, 2014; Nicolle et al., 2012). In the context of the imitation-inhibition task, this relates to identifying which of the representations is most relevant to task performance – the motor representation pertaining to the other or that of the required response of the self. There may be an interesting avenue for teasing apart these two dimensions in the future, however, at present it remains unclear how a

mechanism of task-relevance differentiation may extend to the range of social cognitive abilities investigated to date in the self-other control literature.

8.3.5. Development and distinctions in self-other control. Finally, an area not touched upon in the current body of work, but which is particularly important for future investigation is the emergence of self-other control during development (Steinbeis, 2016). This will not only aid our understanding of typical and atypical social cognitive development, but will also provide a better understanding of how to target interventions towards training self-other control in individuals experiencing atypical social cognitive functioning.

Moreover, by studying the emergence of self-other control during early development, potential distinctions between self-other control in the motor, cognitive and affective domains may be established (Lamm, Bukowski, & Silani, 2015). For example, we know that imitative behaviour has the capacity to emerge very early in development (see Cook, Bird, Catmur, Press, & Heyes, 2014 for a review of the emergence of imitative behaviour), whereas processing emotions and mental states of the self and others requires additional high-level abilities and thus require more time to develop (Bischof-Kohler, 1994; Wellman, Cross, & Watson, 2001). In fact, Steinbeis (2016) in his review of the current literature proposes that self-other control of cognitive and emotional states is already dissociable in early childhood. Moreover, although the TPJ appears to serve an underlying, common process involved in various social domains, there is reason to believe that posterior and anterior regions of the TPJ may also be selective to cognitive and affective domains (Mars et al., 2012; Steinbeis, 2016).

8.6. Conclusion

The studies from the current thesis give an insight into the role of the neurocognitive mechanism of self-other control in typical and atypical social cognition. That is how the

ability to maintain neural representations of the self and others, whether these are pertaining to motor, visual perspective, affective state, or opinion representations, as well as the ability to inhibit or enhance these representations according to the demands of the situation, impact on social cognitive functioning. The current body of work develops past research in this area by demonstrating that self-other control is indeed relevant to overcoming conflict between self and other representations, from low to high level social domains; from imitation to lie detection and social conformity. Novel tasks were successfully developed in this work for use in future studies (Chapters 5 and 7).

Moreover, the first investigations of self-other control in disorders of social cognition were completed, specifically with respect to ASD (Chapters 2, 6 and 7) and alexithymia (Chapter 3). These chapters revealed relationships between individual difference measures of autistic and alexithymic trait severities and the integrity of self-other control. However, interesting observations can be made of the common and dissociable role for this mechanism in the social cognitive impairments experienced across the spectrum of the two conditions considered. The current thesis adds to the current literature regarding the role of the TPJ in this mechanism, whilst demonstrating that its involvement extends to higher level social domains more similar to everyday social encounters and concerning the opinions of the self and other.

Taken together, whether impairments in TPJ function and self-other control integrity are acquired or developmental in origin, which is an important future direction for this research, self-other control emerges as a neurocognitive mechanism capable of explaining both typical and atypical social cognitive functioning. By highlighting a mechanism with the potential to explain many facets of the broad profile of social impairments in autism, as well as other disorders with similar social symptom profiles, researchers may be better equipped to

advise on neurocognitive markers and possible interventions for common disorders of social cognition.

It is important to keep in mind that we cannot presume a single cause or mechanism for the complexities of human social cognition. Nevertheless, investigation of such a low-level mechanism in future research is warranted. In particular, the exploration of the developmental origins of self-other control and how this may account for different developmental trajectories across social cognitive abilities will greatly advance our understanding of this mechanism. Finally, appreciating that no area of the brain works in isolation to support a particular cognitive function, it is also important for research to establish the neural network (centring most likely on the TPJ and mPFC) which governs this unique and vital ability to control competing representations of the self and other.

8.6. References

Ainley, V., Brass, M., & Tsakiris, M. (2014). Heartfelt imitation: High interoceptive awareness is linked to greater automatic imitation. *Neuropsychologia, 60,* 21–28.

Allen, P., Freeman, D., Johns, L., & McGuire, P. (2006). Misattribution of self-generated speech in relation to hallucinatory proneness and delusional ideation in healthy volunteers. *Schizophrenia Research, 84,* 281-288.

Allen, P. P., Johns, L. C., Fu, C. H., Broome, M. R., Vythelingum, G. N., & McGuire, P. K. (2004). Misattribution of external speech in patients with hallucinations and delusions. *Schizophrenia Research, 69,* 277-287.

Allen, P., Larøi, F., McGuire, P. K., & Aleman, A. (2008). The hallucinating brain: A review of structural and functional neuroimaging studies of hallucinations. *Neuroscience & Biobehavioral Reviews, 32,* 175-191.

Ashwood, K. L., Gillan, N., Horder, J., Hayward, H., Woodhouse, E., McEwen, F. S., ... & Cadman, T. (2016). Predicting the diagnosis of autism in adults using the Autism-Spectrum Quotient (AQ) questionnaire. *Psychological Medicine, 46*, 2595-2604.

Baron-Cohen, S., Leslie, A.M., & Frith, U. (1985). Does the autistic child have a ‘‘theory of mind?’’ *Cognition*, *21*, 37–46.

Benedetti, F., Bernasconi, A., Bosia, M., Cavallaro, R., Dallaspezia, S., Falini, A., ... & Smeraldi, E. (2009). Functional and structural brain correlates of theory of mind and empathy deficits in schizophrenia. *Schizophrenia Research, 114,* 154-160.

Bird, G., Catmur, C., Silani, G., Frith, C., & Frith, U. (2006). Attention does not modulate neural responses to social stimuli in autism spectrum disorders. *NeuroImage, 31*, 1614-1624.

Bischof-Köhler, D. (1994). Self object and interpersonal emotions. Identification of own mirror image, empathy and prosocial behavior in the 2nd year of life. *Zeitschrift fur Psychologie mit Zeitschrift fur Angewandte Psychologie*, *202*, 349-377.

Bond, R., & Smith, P. B. (1996). Culture and conformity: a meta-analysis of studies using Asch's (1952b 1956) line judgement task. *Psychological Bulletin*, *119*, 111-137.

Borckardt, J. J., Bikson, M., Frohman, H., Reeves, S. T., Datta, A., Bansal, V., ... & George, M. S. (2012). A pilot study of the tolerability and effects of high-definition transcranial direct current stimulation (HD-tDCS) on pain perception. *The Journal of Pain*, *13*, 112-120.

Borckardt, J. J., Romagnuolo, J., Reeves, S. T., Madan, A., Frohman, H., Beam, W., & George, M. S. (2011). Feasibility, safety, and effectiveness of transcranial direct current stimulation for decreasing post-ERCP pain: A randomized, sham-controlled, pilot study. *Gastrointestinal Endoscopy*, *73*, 1158-1164.

Brass, M., Bekkering, H., & Prinz, W. (2001). Movement observation affects movement execution in a simple response task. *Acta Psychologica*, *106*, 3-22.

Brass, M., Derrfuss, J., Matthes-von Cramon, G., & von Cramon, D. Y. (2003). Imitative response tendencies in patients with frontal brain lesions. *Neuropsychology*, *17*, 265-271.

Brass, M., Derrfuss, J., & von Cramon, D. Y. (2005). The inhibition of imitative and overlearned responses: a functional double dissociation. *Neuropsychologia*, *43*, 89-98.

Brass, M., Ruby, P., & Spengler, S. (2009). Inhibition of imitative behaviour and social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 2359-2367.

Brewer, R., Cook, R., & Bird, G. (2014). Alexithymia explains impaired emotion recognition in eating disorders and schizophrenia. *Journal of Vision, 14*, 1439-1439.

Brewer, R., Cook, R., & Bird, G. (2016a). Shared interoceptive representations: The case of alexithymia. In S. S. Obhi & E. S. Cross (Eds.), *Shared representations: Sensorimotor foundations of social life*. Cambridge, UK: Cambridge University Press.

Brewer, R., Cook, R., & Bird, G. (2016b). Alexithymia: A general deficit of interoception. *Royal Society Open Science, 3*, 150664.

Brewer, R., Cook, R., Cardi, V., Treasure, J., & Bird, G. (2015). Emotion recognition deficits in eating disorders are explained by co-occurring alexithymia. *Royal Society Open Science, 2*, 140382.

Brüne, M., Lissek, S., Fuchs, N., Witthaus, H., Peters, S., Nicolas, V., ... & Tegenthoff, M. (2008). An fMRI study of theory of mind in schizophrenic patients with "passivity" symptoms. *Neuropsychologia, 46*, 1992-2001.

Brüne, M., Özgürdal, S., Ansorge, N., von Reventlow, H. G., Peters, S., Nicolas, V., ... & Lissek, S. (2011). An fMRI study of "theory of mind" in at-risk states of psychosis: Comparison with manifest schizophrenia and healthy controls. *NeuroImage, 55*, 329-337.

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

Catmur, C., Walsh, V., & Heyes, C. (2009). Associative sequence learning: the role of experience in the development of imitation and the mirror system. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*, 2369-2380.

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences, 16,* 231-239.

Cook, J. L. (2014). Task-relevance dependent gradients in medial prefrontal and temporoparietal cortices suggest solutions to paradoxes concerning self/other control. *Neuroscience & Biobehavioral Reviews, 42,* 298-302.

Cook, J., Barbalat, G., & Blakemore, S. J. (2012). Top-down modulation of the perception of other people in schizophrenia and autism. *Frontiers in Human Neuroscience, 6, 175*.

Cook, R., Bird, G., Catmur, C., Press, C., & Heyes, C. (2014). Mirror neurons: From origin to function. *Behavioral and Brain Sciences, 37,* 177-192.

Cook, R., Brewer, R., Shah, P., & Bird, G. (2013). Alexithymia, not autism, predicts poor recognition of emotional facial expressions. *Psychological Science, 24,* 723-732.

Costa, A., Torriero, S., Oliveri, M., & Caltagirone, C. (2008). Prefrontal and temporo-parietal involvement in taking others' perspective: TMS evidence. *Behavioural Neurology, 19,* 71-74.

Courchesne, E., Townsend, J., Akshoomoff, N. A., Saitoh, O., Yeung-Courchesne, R., Lincoln, A. J., ... & Lau, L. (1994). Impairment in shifting attention in autistic and cerebellar patients. *Behavioral Neuroscience, 108,* 848-856.

Crutchfield, R. S. (1955). Conformity and character. *American Psychologist, 10,* 191-198.

Das, P., Lagopoulos, J., Coulston, C. M., Henderson, A. F., & Malhi, G. S. (2012). Mentalizing impairment in schizophrenia: A functional MRI study. *Schizophrenia Research, 134,* 158-164.

de Achával, D., Villarreal, M. F., Costanzo, E. Y., Douer, J., Castro, M. N., Mora, M. C., ... & Guinjoan, S. M. (2012). Decreased activity in right-hemisphere structures involved

in social cognition in siblings discordant for schizophrenia. *Schizophrenia Research, 134*, 171-179.

Demirtas-Tatlidede, A., Vahabzadeh-Hagh, A. M., & Pascual-Leone, A. (2013). Can noninvasive brain stimulation enhance cognition in neuropsychiatric disorders?. *Neuropharmacology, 64*, 566-578.

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., ... & Convit, A. (2006). Introducing MASC: A movie for the assessment of social cognition. *Journal of Autism and Developmental Disorders, 36*, 623-636.

Edwards, D., Cortes, M., Datta, A., Minhas, P., Wassermann, E. M., & Bikson, M. (2013). Physiological and modeling evidence for focal transcranial electrical brain stimulation in humans: A basis for high-definition tDCS. *NeuroImage, 74*, 266-275.

Farrokhyar, F., Reddy, D., Poolman, R. W., & Bhandari, M. (2013). Practical tips for surgical research: Why perform a priori sample size calculation?. *Canadian Journal of Surgery, 56*, 207-213.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Ferri, F., Frassinetti, F., Mastrangelo, F., Salone, A., Ferro, F. M., & Gallese, V. (2012). Bodily self and schizophrenia: The loss of implicit self-body knowledge. *Consciousness and Cognition, 21*, 1365-1374.

Fregni, F., Boggio, P. S., Santos, M. C., Lima, M., Vieira, A. L., Rigonatti, S. P., . . .& Pascual-Leone, A. (2006). Noninvasive cortical stimulation with transcranial direct

current stimulation in parkinson's disease. *Movement Disorders: Official Journal of the Movement Disorder Society, 21*, 1693-702.

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex, 86*, 186-204.

Graham, E. K., & Lachman, M. E. (2014). Personality traits, facets and cognitive performance: Age differences in their relations. *Personality and Individual Differences, 59*, 89-95.

Graham-Schmidt, K. T., Martin-Iverson, M. T., & Waters, F. A. (in press). Self-and other-agency in people with passivity (first rank) symptoms in schizophrenia. *Schizophrenia Research*. doi.org/10.1016/j.schres.2017.04.024.

Hamilton, A. F. D. C. (2013). Reflecting on the mirror neuron system in autism: A systematic review of current theories. *Developmental Cognitive Neuroscience, 3*, 91-105.

Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders, 24*, 129-154.

Happé, F., Booth, R., Charlton, R., & Hughes, C. (2006). Executive function deficits in autism spectrum disorders and attention-deficit/hyperactivity disorder: Examining profiles across domains and ages. *Brain and Cognition, 61*, 25–39.

Happé, F., & Frith, U. (1995). Theory of mind in autism. In E. Schopler & G. B. Mesibov (Eds.), *Learning and cognition in autism* (pp. 177–197). New York, NY: Springer.

Heiser, M., Iacoboni, M., Maeda, F., Marcus, J., & Mazziotta, J. C. (2003). The essential role of Broca's area in imitation. *European Journal of Neuroscience, 17*, 1123-1128.

Hill, E. L. (2004). Evaluating the theory of executive dysfunction in autism. *Developmental Review, 24*, 189-233.

Hogeveen, J., Obhi, S. S., Banissy, M. J., Santiesteban, I., Press, C., Catmur, C., & Bird, G. (2015). Task-dependent and distinct roles of the temporoparietal junction and inferior frontal cortex in the control of imitation. *Social Cognitive and Affective Neuroscience*, *10*, 1003-1009.

Horvath, J. C., Forte, J. D., & Carter, O. (2015a). Quantitative review finds no evidence of cognitive effects in healthy populations from single-session transcranial direct current stimulation (tDCS). *Brain Stimulation*, *8*, 535-550.

Horvath, J. C., Forte, J. D., & Carter, O. (2015b). Evidence that transcranial direct current stimulation (tDCS) generates little-to-no reliable neurophysiologic effect beyond MEP amplitude modulation in healthy human subjects: A systematic review. *Neuropsychologia*, *66*, 213-236.

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

Jeannerod, M. (2009). The sense of agency and its disturbances in schizophrenia: A reappraisal. *Experimental Brain Research*, *192*, 527-532.

Kane, M. J., Meier, M. E., Smeekens, B. A., Gross, G. M., Chun, C. A., Silvia, P. J., & Kwapil, T. R. (2016). Individual differences in the executive control of attention, memory, and thought, and their associations with schizotypy. *Journal of Experimental Psychology: General*, *145*, 1017-1048.

Koeda, M., Takahashi, H., Matsuura, M., Asai, K., & Okubo, Y. (2013). Cerebral responses to vocal attractiveness and auditory hallucinations in schizophrenia: A functional MRI study. *Frontiers in Human Neuroscience*, *7*, 221.

Krall, S. C., Rottschy, C., Oberwelland, E., Bzdok, D., Fox, P. T., Eickhoff, S. B., ... & Konrad, K. (2015). The role of the right temporoparietal junction in attention and social interaction as revealed by ALE meta-analysis. *Brain Structure and Function*, *220*, 587-604.

Lamm, C., Bukowski, H., & Silani, G. (2016). From shared to distinct self-other representations in empathy: Evidence from neurotypical function and socio-cognitive disorders. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*, 20150083.

Langdon, R., Michie, P. T., Ward, P. B., McConaghy, N., Catts, S. V., & Coltheart, M. (1997). Defective self and/or other mentalising in schizophrenia: A cognitive neuropsychological approach. *Cognitive Neuropsychiatry*, *2*, 167-193.

Lee, J., Quintana, J., Nori, P., & Green, M. F. (2011). Theory of mind in schizophrenia: Exploring neural mechanisms of belief attribution. *Social Neuroscience*, *6*, 569-581.

Leekam, S. (2016). Social cognitive impairment and autism: what are we trying to explain?. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*, 20150082.

Lindenberg, R., Renga, V., Zhu, L. L., Nair, D., & Schlaug, G. M. D. P. (2010). Bihemispheric brain stimulation facilitates motor recovery in chronic stroke patients. *Neurology*, *75*, 2176-2184.

Looi, C. Y., & Cohen Kadosh, R. (2016). Brain stimulation, mathematical, and numerical training: Contribution of core and noncore skills. *Progress in Brain Research, 227*, 353-388.

Mars, R. B., Sallet, J., Schüffelgen, U., Jbabdi, S., Toni, I., & Rushworth, M. F. (2012). Connectivity-based subdivisions of the human right "temporoparietal junction area": Evidence for different areas participating in different cortical networks. *Cerebral Cortex*, *22*, 1894-1903.

Mechelli, A., Allen, P., Amaro, E., Fu, C. H., Williams, S. C., Brammer, M. J., ... & McGuire, P. K. (2007). Misattribution of speech and impaired connectivity in patients with auditory verbal hallucinations. *Human Brain Mapping*, *28*, 1213-1222.

Morrier, M. J., Ousley, O. Y., Caceres-Gamundi, G. A., Segall, M. J., Cubells, J. F., Young, L. J., & Andari, E. (2017). Relationship between ADOS-2, module 4 calibrated severity scores (CSS) and social and non-social standardized assessment measures in adult males with autism spectrum disorder (ASD). *Journal of Autism and Developmental Disorders*. doi.org/10.1007/s10803-017-3293-z.

Murphy, J., Brewer, R., Catmur, C., & Bird, G. (2017). Interoception and psychopathology: A developmental neuroscience perspective. *Developmental Cognitive Neuroscience*, *23*, 45-56.

Murphy, J., Catmur, C., & Bird, G. (in press). Alexithymia is associated with a multi-domain, multi-dimensional failure of interoception: Evidence from novel tests. *Journal of Experimental Psychology*: *General*.

Nicolle, A., Klein-Flügge, M. C., Hunt, L. T., Vlaev, I., Dolan, R. J., & Behrens, T. E. (2012). An agent independent axis for executed and modeled choice in medial prefrontal cortex. *Neuron*, *75*, 1114-1121.

Nitsche, M. A., Boggio, P. S., Fregni, F., & Pascual-Leone, A. (2009). Treatment of depression with transcranial direct current stimulation (tDCS): A review. *Experimental Neurology, 219,* 14-19.

Ozonoff, S. (1997). Components of executive function in autism and other disorders. In J. Russell (Ed.), *Autism as an executive disorder* (pp. 179–211). New York, NY: Oxford University Press.

Parkin, B. L., Ekhtiari, H., & Walsh, V. F. (2015). Non-invasive human brain stimulation in cognitive neuroscience: A primer. *Neuron, 87,* 932-945.

Price, A. R., & Hamilton, R. H. (2015). A re-evaluation of the cognitive effects from single-session transcranial direct current stimulation. *Brain Stimulation, 8,* 663-665.

Quattrocki, E., & Friston, K. (2014). Autism, oxytocin and interoception. *Neuroscience & Biobehavioral Reviews, 47,* 410–430.

Ramachandran, V. S., & Oberman, L. M. (2006). Broken mirrors: A theory of autism. *Scientific American, 295,* 62-69.

Reis, J., Schambra, H. M., Cohen, L. G., Buch, E. R., Fritsch, B., Zarahn, E., ... & Krakauer, J. W. (2009). Noninvasive cortical stimulation enhances motor skill acquisition over multiple days through an effect on consolidation. *Proceedings of the National Academy of Sciences, 106,* 1590-1595.

Renes, R. A., Vermeulen, L., Kahn, R. S., Aarts, H., & van Haren, N. E. (2013). Abnormalities in the establishment of feeling of self-agency in schizophrenia. *Schizophrenia Research, 143,* 50-54.

Roe, D., & Davidson, L. (2005). Self and narrative in schizophrenia: Time to author a new story. *Medical Humanities, 31,* 89-94.

Russell, J. (1997). *Autism as an executive disorder*. New York, NY: Oxford University Press.

Russo, R., Wallace, D., Fitzgerald, P. B., & Cooper, N. R. (2013). Perception of comfort during active and sham transcranial direct current stimulation: A double blind study. *Brain Stimulation, 6,* 946-951.

Santiesteban, I., Banissy, M. J., Catmur, C., & Bird, G. (2012a). Enhancing social ability by stimulating right temporoparietal junction. *Current Biology, 22,* 2274-2277.

Santiesteban, I., Banissy, M. J., Catmur, C., & Bird, G. (2015). Functional lateralization of temporoparietal junction–imitation inhibition, visual perspective- taking and theory of mind. *European Journal of Neuroscience*, *42*, 2527-2533.

Santiesteban, I., White, S., Cook, J., Gilbert, S. J., Heyes, C., & Bird, G. (2012b). Training social cognition: From imitation to theory of mind. *Cognition*, *122*, 228-235.

Saxe, R. (2009). Theory of mind (neural basis). *Encyclopedia of Consciousness, 2*, 401-410.

Saxe, R. (2010). The right temporo-parietal junction: A specific brain region for thinking about thoughts. In A. Leslie & T. German (Eds.), *Handbook of theory of mind*. Hove, UK: Psychology Press.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind". *NeuroImage, 19*, 1835-1842.

Saxe, R., & Wexler, A. (2005). Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia, 43*, 1391-1399.

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.

Sellaro, R., Nitsche, M. A., & Colzato, L. S. (2017). Transcranial direct current stimulation. In L. S. Colzato (Ed.), *Theory-driven approaches to cognitive enhancement* (pp. 99-112). Switzerland, CH: Springer International Publishing.

Senju, A., & Johnson, M. H. (2009). Atypical eye contact in autism: Models, mechanisms and development. *Neuroscience & Biobehavioral Reviews*, *33*, 1204-1214.

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences, 17,* 565–573.

Shah, P., Hall, R., Catmur, C., & Bird, G. (2016). Alexithymia, not autism, is associated with impaired interoception. *Cortex, 81,* 215–220.

Shiozawa, P., Fregni, F., Benseñor, I. M., Lotufo, P. A., Berlim, M. T., Daskalakis, J. Z., ... & Brunoni, A. R. (2014). Transcranial direct current stimulation for major depression: An updated systematic review and meta-analysis. *International Journal of Neuropsychopharmacology*, *17*, 1443-1452.

Sowden, S., Brewer, R., Catmur, C., & Bird, G. (2016b). The specificity of the link between alexithymia, interoception, and imitation. *Journal of Experimental Psychology: Human Perception and Performance*, *42*, 1687-1692.

Sowden, S., & Catmur, C. (2015). The role of the right temporoparietal junction in the control of imitation. *Cerebral Cortex*, *25*, 1107-1113.

Sowden, S., Koehne, S., Catmur, C., Dziobek, I., & Bird, G. (2016a). Intact automatic imitation and typical spatial compatibility in autism spectrum disorder: Challenging the broken mirror theory. *Autism Research*, *9*, 292-300.

Sowden, S., & Shah, P. (2014). Self-other control: A candidate mechanism for social cognitive function. *Frontiers in Human Neuroscience*, *8*, 789.

Sowden, S., Wright, G. R., Banissy, M. J., Catmur, C., & Bird, G. (2015). Transcranial current stimulation of the temporoparietal junction improves lie detection. *Current Biology*, *25*, 2447-2451.

Spengler, S., von Cramon, D. Y., & Brass, M. (2009a). Control of shared representations relies on key processes involved in mental state attribution. *Human Brain Mapping*, *30*, 3704-3718.

Spengler, S., von Cramon, D. Y., & Brass, M. (2009b). Was it me or was it you? How the sense of agency originates from ideomotor learning revealed by fMRI. *NeuroImage*, *46*, 290-298.

Spengler, S., von Cramon, D. Y., & Brass, M. (2010). Resisting motor mimicry: Control of imitation involves processes central to social cognition in patients with frontal and temporo-parietal lesions. *Social Neuroscience, 5*, 401-416.

Steinbeis, N. (2016). The role of self-other distinction in understanding others' mental and emotional states: Neurocognitive mechanisms in children and adults. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*, 20150074.

Stürmer, B., Aschersleben, G., & Prinz, W. (2000). Correspondence effects with manual gestures and postures: A study of imitation. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 1746–1759.

Tandon, R., Keshavan, M. S., & Nasrallah, H. A. (2008). Schizophrenia, "just the facts": what we know in 2008: Part 1: Overview. *Schizophrenia Research, 100*, 4-19.

Torrey, E. F. (2007). Schizophrenia and the inferior parietal lobule. *Schizophrenia Research*, *97*, 215-225.

van der Meer, L., Costafreda, S., Aleman, A., & David, A. S. (2010). Self-reflection and the brain: a theoretical review and meta-analysis of neuroimaging studies with implications for schizophrenia. *Neuroscience & Biobehavioral Reviews, 34,* 935-946.

van der Weiden, A., Prikken, M., & van Haren, N. E. (2015). Self–other integration and distinction in schizophrenia: A theoretical analysis and a review of the evidence. *Neuroscience & Biobehavioral Reviews, 57,* 220-237.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta- analysis of theory- of- mind development: The truth about false belief. *Child Development, 72,* 655-684.

Williams, D.L., Goldstein, G., Carpenter, P.A., & Minshew, N.J. (2005). Verbal and spatial working memory in autism. *Journal of Autism and Developmental Disorders, 35,* 747–756.

Williams, J.H., Whiten, A., & Singh, T. (2004). A systematic review of action imitation in autistic spectrum disorder. *Journal of Autism and Developmental Disorders, 34,* 285–299.

Appendices

Appendix A: Opinion Questionnaire

Please answer the following questions as honestly as you are able. You are asked to give your personal stance on the topic presented. The scale goes from 1 - 6 with 1 indicating that you are "Strongly for" the topic and 6 indicating that you are "Strongly against" the topic.

1. Medical Abortion	Strongly for Strongly against 1 2 3 4 5 6
2. Euthanasia	Strongly for Strongly against 1 2 3 4 5 6
3. Horoscopes	Strongly for Strongly against 1 2 3 4 5 6
4. The idea of Climate Change	Strongly for Strongly against 1 2 3 4 5 6
5. Gay marriage	Strongly for Strongly against 1 2 3 4 5 6
6. Genetically Modified Foodstuffs	Strongly for Strongly against 1 2 3 4 5 6
7. Unemployment benefits	Strongly for Strongly against 1 2 3 4 5 6
8. Animal Testing	Strongly for Strongly against 1 2 3 4 5 6
9. Nuclear power	Strongly for Strongly against 1 2 3 4 5 6
10. Pornography	Strongly for Strongly against 1 2 3 4 5 6
11. Homeopathic medicine	Strongly for Strongly against 1 2 3 4 5 6
12. Lying	Strongly for Strongly against 1 2 3 4 5 6
13. Free University	Strongly for Strongly against 1 2 3 4 5 6
14. School Uniform	Strongly for Strongly against 1 2 3 4 5 6
15. Stem Cell research	Strongly for Strongly against 1 2 3 4 5 6
16. The Monarchy	Strongly for Strongly against 1 2 3 4 5 6
17. Television	Strongly for Strongly against 1 2 3 4 5 6
18. Alcohol	Strongly for Strongly against 1 2 3 4 5 6
19. Cosmetic / Plastic Surgery	Strongly for Strongly against 1 2 3 4 5 6
20. Videogames	Strongly for Strongly against 1 2 3 4 5 6

Appendix B: Autism Spectrum Quotient

Read each statement and tick the box to the right of the statement that corresponds best to your answer. There are no right or wrong answers. Do not spend too much time on any one statement.	Definitely agree	Slightly agree	Slightly disagree	Definitely disagree
1. I prefer to do things with others rather than on my own.				
2. I prefer to do things the same way over and over again.				
3. If I try to imagine something I find it very easy to create a picture in my mind.				
4. I frequently get so absorbed in one thing that I lose sight of other things.				
5. I often notice small sounds when others do not.				
6. I usually notice car number plates or similar strings of information.				
7. Other people frequently tell me that what I've said is impolite, even though I think it is polite.				
8. When I'm reading a story, I can easily imagine what the characters might look like.				
9. I am fascinated by dates.				
10. In a social group, I can easily keep track of several different people's conversations.				
11. I find social situations easy.				
12. I tend to notice details that others do not.				
13. I would rather go to a library than a party.				
14. I find making up stories easy.				
15. I find myself drawn more strongly to people than to things.				
16. I tend to have very strong interests, which I get upset about if I can't pursue.				
17. I enjoy social chit-chat.				
18. When I talk, it isn't always easy for others to get a word in edgeways.				
19. I am fascinated by numbers.				
20. When I'm reading a story, I find it difficult to work out the characters' intentions.				

21. I don't particularly enjoy reading fiction.				
22. I find it hard to make new friends.				
23. I notice patterns in things all the time.				
24. I would rather go to the theatre than to a museum.				
25. It does not upset me if my daily routine is disturbed.				
26. I frequently find that I don't know how to keep a conversation going.				
27. I find it easy to « read between the lines » when someone is talking to me.				
28. I usually concentrate more on the whole picture rather than the small details.				
29. I'm not very good at remembering phone numbers.				
30. I don't usually notice small changes in a situation, or a person's appearance.				
31. I know how to tell if someone listening to me is getting bored.				
32. I find it easy to do more than one thing at once.				
33. When I talk on the phone, I'm not sure when its my turn to speak.				
34. I enjoy doing things spontaneously.				
35. I am often the last to understand the point of a joke.				
36. I find it easy to work out what someone is thinking or feeling just by looking at their face.				
37. If there is an interruption, I can switch back to what I was doing very quickly.				
38. I am good at social chit-chat.				
39. People often tell me that I keep going on and on about the same thing.				
40. When I was young, I used to enjoy playing games involving pretending with other children.				
41. I like to collect information about categories of things (e.g. types of car, types of bird, types of train, types of plant etc)				
42. I find it difficult to imagine what it would be like to be someone else.				
43. I like to plan any activities I participate in carefully.				

44. I enjoy social occasions.				
45. I find it difficult to work out people's intentions.				
46. New situations make me anxious.				
47. I enjoy meeting new people.				
48. I am a good diplomat.				
49. I am not very good at remembering people's date of birth.				
50. I find it very easy to play games with children that involve pretending.				

Appendix C: Toronto Alexithymia Scale

Indicate how much you agree or disagree with each of the following statements. Use the middle answer ('I neither agree or disagree') only if you are really unable to assess your behaviour. When you have finished, please inform the experimenter.

1. I am often confused about what emotion I am feeling	Strongly Disagree 1 2 3 4 5 Strongly Agree
2. It is difficult for me to find the right words for my feelings	Strongly Disagree 1 2 3 4 5 Strongly Agree
3. I have physical sensations that even doctors don't understand	Strongly Disagree 1 2 3 4 5 Strongly Agree
4. I am able to describe my feelings easily	Strongly Disagree 1 2 3 4 5 Strongly Agree
5. I prefer to analyze problems rather than just describe them	Strongly Disagree 1 2 3 4 5 Strongly Agree
6. When I am upset, I don't know if I am sad, frightened, or angry	Strongly Disagree 1 2 3 4 5 Strongly Agree
7. I am often puzzled by sensations in my body	Strongly Disagree 1 2 3 4 5 Strongly Agree
8. I prefer to just let things happen rather than to understand why they turned out that way	Strongly Disagree 1 2 3 4 5 Strongly Agree
9. I have feelings that I can't quite identify	Strongly Disagree 1 2 3 4 5 Strongly Agree
10. Being in touch with emotions is essential	Strongly Disagree 1 2 3 4 5 Strongly Agree
11. I find it hard to describe how I feel about people	Strongly Disagree 1 2 3 4 5 Strongly Agree
12. People tell me to describe my feelings more	Strongly Disagree 1 2 3 4 5 Strongly Agree
13. I often don't know what's going on inside me	Strongly Disagree 1 2 3 4 5 Strongly Agree
14. I often don't know why I am angry	Strongly Disagree 1 2 3 4 5 Strongly Agree
15. I prefer talking to people about their daily activities rather than their feelings	Strongly Disagree 1 2 3 4 5 Strongly Agree
16. I prefer to watch light entertainments shows rather than psychological dramas	Strongly Disagree 1 2 3 4 5 Strongly Agree
17. It is difficult for me to reveal my innermost feelings, even to close friends	Strongly Disagree 1 2 3 4 5 Strongly Agree
18. I can feel close to someone, even in moments of silence	Strongly Disagree 1 2 3 4 5 Strongly Agree
19. I find examination of my feelings useful in solving personal problems	Strongly Disagree 1 2 3 4 5 Strongly Agree
20. Looking for hidden meanings in movies or plays distracts from their enjoyment	Strongly Disagree 1 2 3 4 5 Strongly Agree

Appendix D: Autism Spectrum Quotient (10 item scale)

Read each statement and tick the box to the right of the statement that corresponds best to your answer. There are no right or wrong answers. Do not spend too much time on any one statement.	Definitely agree	Slightly agree	Slightly disagree	Definitely disagree
1. I often notice small sounds when others do not				
2. I usually concentrate more on the whole picture, rather than the small details				
3. I find it easy to do more than one thing at once				
4. If there is an interruption, I can switch back to what I was doing very quickly				
5. I find it easy to 'read between the lines' when someone is talking to me				
6. I know how to tell if someone listening to me is getting bored				
7. When I'm reading a story I find it difficult to work out the characters' intentions				
8. I like to collect information about categories of things (e.g. types of car, types of bird, types of train, types of plant etc.)				
9. I find it easy to work out what someone is thinking or feeling just by looking at their face				
10. I find it difficult to work out people's intentions				

SCORING: Only 1 point can be scored for each question. *Score 1 point for Definitely or Slightly Agree on each of items 1, 7, 8, and 10. Score 1 point for Definitely or Slightly Disagree on each of items 2, 3, 4, 5, 6, and 9.* If the individual scores **more than 6 out of 10**, consider referring them for a specialist diagnostic assessment.