DOCTOR OF PHILOSOPHY

An fMRI Investigation of Visual Responses to Social Interactions

Walbrin, Jonathan

Award date:
2019

Awarding institution:
Bangor University

An fMRI Investigation of Visual Responses to Social Interactions

Jon Walbrin

Bangor University

PhD Psychology

2019

For my long-suffering parents, Michael and Pauline

PhD Committee

Kami Koldewyn (first supervisor)

Paul Downing (second supervisor)

Manon Jones (chair)

Declaration

I hereby declare that this thesis is the result of my own investigations, except where otherwise stated. All other sources are acknowledged by bibliographic references. This work has not previously been accepted in substance for any other degree and is not being concurrently submitted in candidature for any degree unless, as agreed by the University, for approved dual awards.

Yr wyf drwy hyn yn datgan mai canlyniad fy ymchwil fy hun yw'r thesis hwn, ac eithrio lle nodir yn wahanol. Caiff ffynonellau eraill eu cydnabod gan droednodiadau yn rhoi cyfeiriadau eglur. Nid yw sylwedd y gwaith hwn wedi cael ei dderbyn o'r blaen ar gyfer unrhyw radd, ac nid yw'n cael ei gyflwyno ar yr un pryd mewn ymgeisiaeth am unrhyw radd oni bai ei fod, fel y cytunwyd gan y Brifysgol, am gymwysterau deuol cymeradwy.

Acknowledgements

Kami, the last four years have been incredible and I'm very thankful to you for offering me this opportunity. I didn't think I would ever feel like a researcher, but your continual encouragement has allowed me to develop the confidence and abilities to complete this PhD. I'm also very grateful for: Your quick and intuitive understanding of good research ideas; remaining approachable and easy-going; nudging me towards opportunities that I would have otherwise missed; your extreme generosity; and your unending patience.

~

To Paul, thank you for your guidance, enthusiasm, and support over the course of the PhD, and keeping me on track.

~

To Susanne and David, thank you for taking the time to read this thesis and making the viva such a rewarding experience.

~

To Simon Watt, Julie Davies, Paloma Mari-Beffa, Zeynep Saygin, and David Osher, thank you for your crucial contributions in the run-up to, and during, this PhD.

~

To Dilini and Faisal, thank you for the unending hilarity, generosity, and incredible food.

~

To Conny, Emma, Ioana, the two Marcos, Julia, Ciaran, Aoife, and Ciara, thank you so much for being great people and for your karaoke compliance.

~

To Michael Gerstenberger, Paul Schulz, Julia Landsiedel, and Ioana Mihai, thank you for your vital contributions to these projects – it was a privilege to work with you.

Summary

Visual 'person perception' research has largely focused on responses to individual human stimuli, for example, single faces or bodies presented in isolation. However, social interactions contain unique social information that cannot be inferred from isolated individuals, for example, whether two interactors share a common goal. Similarly, recent behavioural research demonstrates that visual responses to face and body information of interacting individuals are qualitatively different than when they are not interacting. However, little is known about the neural basis of these responses to interactive behaviour.

The work presented in this thesis aims to characterize the neural correlates of third person social interaction perception. Across three functional magnetic resonance imaging (fMRI) studies, converging evidence implicates the posterior temporal lobe as a key region in recognizing, differentiating, and integrating interactive information. The results from these experiments suggest that the posterior superior temporal sulcus (pSTS) and extrastriate body area (EBA) play important complementary roles in third person social interaction perception, making valuable contributions to the emerging field of social interaction research.

Contents

“I tell you, we are here on Earth to fart around,
and don't let anybody tell you different.”

Kurt Vonnegut

Chapter 1

General Introduction

1.1. Rationale for this Thesis

Social interactions are ubiquitous social scenarios that convey unique social information, such as the shared goals, intentions, or the relative social statuses of interactors. Although, precise definitions of what exactly constitutes a social interaction might vary, De Jaeger, Di Paolo, and Gallagher (2010) present the following precise yet concise definition: "…two or more autonomous agents co-regulating their coupling with the effect that their autonomy is not destroyed and their relational dynamics acquire an autonomy of their own." This account contains two important characteristics of social interactions that are paraphrased as follows: Firstly, that interacting individuals are not only autonomous entities, but also contribute to a shared interactive context; and secondly, that the behaviour or actions of each interactor dynamically alters the behaviour of other interactors.

Given these crucial aspects of social interactions, it is clear that traditional person perception research is insufficient to fully explain the perception of multiple interacting individuals. For example, such research has typically focused on responses to individual humans, in isolation, and focused largely on responses to faces (e.g. Duchaine & Yovel, 2015) and bodies (e.g. Peelen & Downing, 2007). However, 'individual' stimuli may not capture certain aspects of social perception that arise from the presence of multiple individuals in the context of a social interaction. For example, configural inversion effects (i.e. quicker and more accurate categorization of single upright bodies and faces, than inverted versions of these stimuli, is taken as a measure of preserved configural processing for upright, but not inverted stimuli; Reed, Stone, Bozova, & Tanaka, 2003) are also observed for pairs of bodies that are configured in an interaction-congruent manner; that is, when two bodies are *facing towards* each other, but crucially, not when *facing away* from each other (Papeo, Stein, & Soto-Faraco, 2017). Similarly, interaction context has been shown to modulate face perception (i.e. emotion judgements of target faces are modulated by the apparent facial expression of a second non-target face when facing towards – but not when facing away from – the target face; Gray, Barber, Murphy, & Cook, 2017). These effects demonstrate the need for a principled investigation of the perception of multi-agent interactive behaviour.

There are numerous possible approaches to studying observational responses to social interactions – such as from either a second- or third-person perspective, or either

with the viewer as an 'active' interactor, or a 'passive' observer (Pfeiffer, Timmermans, Vogeley, Frith, & Schilbach, 2013). Additionally, widely varying stimuli and task instructions across studies (Quadflieg & Koldewyn, 2017), and the frequently noted lack of empirical social interaction perception literature (Gray et al., 2017; Papeo et al., 2017; Quadflieg & Koldewyn, 2017) have meant that few conclusions can be drawn about interaction perception relative to individual person perception literature. Furthermore, social interactions are complex scenarios and fully understanding them likely relies on three different appraisals of information: Perception, action, and social based inferences about interactors. For example, an observer may notice that two interactors are smiling together or share a similar physical resemblance to each other (perceptual appraisal), or that the they are performing complementary or reciprocal movements and behaviours (action appraisal) or may make higher level attributions about the relative social statuses of the two interactors (social appraisal).

Specifically, this thesis will investigate the neural correlates underlying the first two aspects outlined here; that is, which brain regions play a central role in the visual processing of *perceptual* and *action* aspects of dynamic interactions with a particular focus on the detection of the presence of an interaction, and the contents of interactions (e.g. whether interactors are competing or cooperating with each other). Before describing the undertaken research, an overview of relevant literature is presented. First, a general overview of the functions associated with 'social' brain networks is provided before addressing existent visual third-person social interaction perception research. A brief description of developmental social person perception is also included to provide theoretical context for the experiment conducted in chapter 4. Finally, this section is concluded with an overview of this thesis and the specific research undertaken.

1.2. Overview of Social Perception, Action, and Cognition in the Brain

Visual person perception has typically been associated with a network of occipitotemporal brain regions that are suggested to play a specialised role in processing specific categories of social information. For example, face *selective* regions such as the fusiform face area (FFA) and occipital face area (OFA), along with body selective regions such as the extrastriate body area (EBA) and fusiform body area (FBA), are shown to be maximally responsive to the presence of their 'preferred' categories of social information

(e.g. Downing, Chan, Peelen, Dodds, & Kanwisher, 2006). Whether these selective responses constitute discrete 'modular' cortical regions, or instead key nodes that interface between networks that are functionally tuned towards physical properties of a given stimulus category of interest, is currently uncertain (Op de Beeck, Haushofer, & Kanwisher, 2008). For example, evidence that individuals with prosopagnosia (i.e. 'face-blindness') show aberrant functional responses in ventral occipital face selective regions (Hadjikhani & de Gelder, 2002) is countered with evidence that these perceptual deficits arise from weaker responses in other regions within a wider face perception network (Avidan & Behrmann, 2009).

Beyond the visual perception of face and body information, the actions and goals of individuals must also be interpreted and understood. Action understanding is a complex process that likely consists of several hierarchical levels (Hamilton & Grafton, 2006). For example, observing and understanding an individual's intended goal – such as pouring a glass of whisky for a friend – may depend on the recognition and processing of several preceding steps such as: 1) The individual's specific arm *movements* (e.g. extending an arm and adjusting a hand, ready to grip); 2) a coherent *action* (i.e. reaching to pick up a bottle), and; 3) an *immediate goal* (e.g. lifting the bottle to pour). Although action understanding requires (typically visual) perceptual information, it is clear that further inferential processing is also required, that might rely on regions that may (or may not) also demonstrate non-visual functional properties. A network of regions often referred to as the 'action observation network' (AON) or 'mirror network' are often implicated in such processing.

Since the discovery of so-called *mirror neurons* in the macaque brain (i.e. neurons that show highly similar F5/inferior frontal gyrus (IFG) spike rates when both observing and enacting a given action; Gallese, Fadiga, Fogassi, & Rizzolatti, 1996), efforts have been made to determine the existence of mirror mechanisms in humans. To this end, the AON has two key nodes in inferior parietal lobule (IPL; Van Overwalle & Baertens, 2009) and IFG, including ventral pre-motor cortex (vPMC); these regions are suggested to play a key role in the translation of sensory representations (i.e. observed visual or auditory actions) into motor representations of the same actions, whilst other regions (e.g. insula and anterior cingulate) are implicated in translating observed emotional content into embodied sensations (Fabbri-Destro & Rizzolatti, 2008). Accordingly, by simulating and translating externally observed actions into the observer's own motor representations,

such mirror mechanisms allegedly allow relatively immediate and non-deliberative understanding of an observed action and the underlying intention behind it (Caspers, Zilles, Laird, & Eickhoff, 2010).

However, such mirror simulation accounts, especially regarding IFG regions, are challenged by findings that suggest responses in this region simply reflect the individual's own first-person motor representations of actions, rather than a 'mirror-matching' response. Oosterhof, Tipper, and Downing (2012) found that neural responses in vPMC when performing hand-based actions generalised to responses when observing similar actions from a *first-person perspective*, but not when viewing actions from a *third-person perspective*. Similarly, Papeo, Corradi-Dell'Acqua, and Rumiati (2011) observed responses consistent with the enhancement of internal motor responses for first-person – but not third-person – hand action verbs (i.e. transcranial magnetic stimulation (TMS) of primary motor cortex induced increases in motor-evoked potentials). These findings suggest that pre-motor/primary motor cortex responses likely reflect first-person motor representations, rather than mirror-matching of another individual's movements.

By contrast, Oosterhof et al. (2012) observed shared representations between the performance of actions and observation of the same action from a third-person perspective in posterior superior parietal cortex. Similarly, Papeo & Lingnau (2015) found that third-person verbs evoked greater responses in posterior temporal cortex. It therefore seems plausible that observation of third-person social interactions may rely more strongly on posterior temporo-parietal regions of the brain, rather than IFG motor cortex.

The relative immediacy of action understanding that is proposed for the AON (Caspers et al., 2010) is complemented by the capacity for higher level inference of others mental-states. Accordingly, so-called 'mentalizing' processes are subserved by a distinct set of core mentalizing network regions – the medial prefrontal cortex (mPFC), paracingulate cortex/precuneus (PCC), and the temporo-parietal junction (TPJ; Van Overwalle & Baertens, 2009). Crucially, these mentalizing regions are typically recruited across a variety of explicit judgement tasks about others' 'hidden' mental states, such as when deliberatively inferring an individual's thoughts or beliefs (Dodell-Feder, Koster-Hale, Bedny, & Saxe, 2011), reasoning about why an individual has performed a given action (Spunt & Adolphs, 2014), or inferring the apparent intentions of moving animate shapes (Castelli, Happé, Frith, & Frith, 2000). Although specific mentalizing definitions

may vary, most researchers agree on a general dichotomy of social inference making – that is, that transitory inferences (e.g. understanding the intentions and goals of an action) are more closely associated with observable, perceptual information, whereas long lasting attributions (e.g. personality traits or long term goal attributions) are suggested to rely more on abstract, mentalizing processing (Van Overwalle, 2009). Similarly, several theories suggest that understanding the immediate intentions that underlie a given action do not require deliberative top-down cognition (e.g. Baron Cohen, 1995; Gallagher, 2004).

Interestingly, an extensive meta-analysis of mentalizing studies revealed that tasks involving inferences derived from dynamic visual action information – for example, inferences based on others' eye gaze, body actions, or moving shapes – tend to recruit regions of the posterior temporal and temporoparietal cortex, while tasks that do not require inferences from visual action observation (e.g. inferring others mental states from written descriptions, or while playing strategic games with another person out of view) strongly recruit mPFC (Schurz, Radua, Aichhorn, Richlan, & Perner, 2014).

The emphasis on posterior temporal regions for visual processing during mentalizing (Gobbini, Koralek, Bryan, Montgomery, & Haxby, 2007; Schurz et al., 2014) and action observation (Cross, Hamilton, Kraemer, Kelley, & Grafton, 2009; Iacoboni 2009; Van Overwalle & Baertens, 2009) suggests that this broad region may play a key role in understanding actions, and potentially, interactions. More specifically, regions of the superior temporal sulcus (STS) are implicated in *intermediate* processes between simple observation of actions and explicit reasoning about another individual's thoughts and beliefs. For example, the STS is sensitive to the *immediate* intentions underlying observed actions (Pelphrey, Morris, & McCarthy, 2004; Shultz, Lee, Pelphrey, & McCarthy, 2010**;** Vander Wyk, Hudac, Carter, Sobel, & Pelphrey, 2009). This is further illustrated by Saxe, Xiao, Kovacs, Perrett, & Kanwisher (2004). Subjects observed brief clips of an individual walking across a room and passing behind – and were therefore temporarily occluded by – a bookshelf. Greater posterior STS (pSTS) responses were found for longer occlusion periods, due to the perceived change in intentional state of the person (e.g. pausing behind the bookshelf after noticing a book). Similarly, the STS is sensitive to the apparent rationality of observed actions (Deen & Saxe, 2012). For example, greater pSTS responses were observed when viewing an individual that walked past a light switch and

used their knee to flip it while their hands were free, compared to when they were carrying a stack of books (Brass, Schmitt, Spengler, & Gergely, 2007).

Complementary to action-intention understanding in the STS, recent evidence shows that a neighbouring region – the lateral occipito-temporal cortex (LOTC) – is sensitive to a wide range of action contents, including visual observation of actions, tool use, and semantic contents of actions (Lingnau & Downing, 2015). This broad region also differentiates between specific actions as well as different action categories (Wurm, Caramazza, & Lingnau, 2017). It is also sensitive to different action categories in an interactive context – that is, when performed by an individual on another individual (e.g. biting or pushing; Hafri, Trueswell, & Epstein, 2017). The relative proximity and potential complementarity of the STS and LOTC suggests that, together, they may play an important role in social interaction perception.

In summary, across these findings, the posterior temporal cortex shows several characteristics that suggest it may play an important role in the visual processing of social interactions; namely, sensitivity to face and body information, action and action-intention understanding. The next section provides an overview of the social functions of the STS – a key region of interest in the empirical chapters of this thesis.

1.3. Overview of the Social Functions of the STS

The STS is a broad region that is sensitive to a variety of social stimuli, that are functionally organized in a complex fashion. Deen, Koldewyn, Kanwisher, and Saxe (2015) observed partially overlapping STS regions that respond strongly to different categories of social information (e.g. faces, theory of mind, voices, and biological motion) as well as regions that respond to multiple social categories of information; subsequent functional connectivity analyses showed that these differential response profiles within the STS are supported by differential coactivation profiles with other brain regions. Similarly, Lahnakoski et al. (2012) propose that the pSTS constitutes a central 'hub' for social information processing that differentially coactivates with different functional networks depending on the specific social stimulus; subjects viewed video clips depicting eight types of social stimuli (i.e. faces, bodies, biological motion, goal-oriented action, emotion, *social interaction,* pain, and speech) and six types of non-social stimuli (houses, objects, rigid motion, non-goal-oriented action, *non-interacting humans*, and non-human

sounds). The bilateral pSTS along with the right fusiform gyrus were the only regions that responded to all social stimuli, and when contrasting all social stimuli (together) against all non-social stimuli, peak responses were observed in the bilateral STS. Functional connectivity analyses also showed the pSTS constitutes a core node in a sub-network of regions that shows task-dependent coactivation with wider brain regions (e.g. face and body responses show differential network coactivations). The authors concluded that the pSTS constitutes a social perception hub; indeed, recent work has also demonstrated similar evidence consistent with this interpretation (Dasgupta, Tyler, Wicks, Srinivasan, & Grossman, 2017).

Converging evidence also suggests that the STS is highly sensitive to 'social motion', including biological motion (i.e. movement of eyes, mouth, hands, body; Allison, Puce, & McCarthy, 2000), as well as strong impressions of intentionality (Gao, Scholl, & McCarthy, 2012) and rationality (Deen & Saxe, 2012) conveyed by the nuanced movements of animate geometric shapes. However, STS regions are also sensitive to static images that imply motion (Allison et al., 2000; Peuskens, Vanrie, Verfaillie, & Orban, 2005) or human actions (Baldassano, Beck, & Fei-Fei, 2017), and similar sensitivity to static implied movements and actual body movement is observed in STS neurons in macaque monkeys (Jellema & Perrett, 2003; Perrett et al., 1985).

In addition to visual social information, evidence of STS regions that are strongly sensitive to human voices (e.g. Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Kriegstein & Giraud, 2004) demonstrate that the wider STS is sensitive to dynamic social information across visual and auditory modalities. Indeed, regions of the STS are also sensitive to both auditory and visual social information and show responses that are consistent with the integration of speech and mouth movements (Calvert, Campbell, & Brammer, 2000), as well as visual and auditory representations of human-object actions (Barraclough, Xiao, Baker, Oram, & Perrett, 2005).

Along with the sensory processing of social stimuli, regions of the STS (including neighbouring TPJ cortex) are also sensitive to higher level 'non-sensory' processing of social information, such as mental state inferences probed by written stories (e.g. Dodell-Feder et al., 2011), social causality judgements between moving geometric shapes (Wende et al., 2013), and explicit social inferences attributed to randomly moving shapes (Lee, Gao, & McCarthy, 2012). One possibility is that the STS may broadly serve as a region that interfaces dynamic sensory and (non-sensory) inferential social information. As

previously mentioned, a host of studies demonstrate that posterior STS is sensitive to the intentions that underlie human actions (e.g. Brass et al., 2007; Pelphrey et al., 2004; Shultz et al., 2010), and as such, this region may play a crucial role in integrating sensory and intentional social information to readily allow recognition and understanding of dynamic social interactions. Indeed, evidence for multimodal integration, and the suggestion that the pSTS serves as a point of convergence for social processing via wider brain networks, highlight the multifaceted and integrative properties of the STS. Whether it plays a further role in integrating information between multiple individuals remains to be seen, but together, these findings implicate the STS as a potentially vital region for the perception and understanding of social interactions.

1.4. Behavioural Perception of Social Interactions

As previously mentioned, configural inversion effects have been observed for 'specialised' social object categories – that is, faces and bodies. Interestingly, similar effects with 'interacting' dyadic body stimuli suggest that social interactions might constitute a specialised social category too. In a series of experiments, Papeo et al. (2017) observed an inversion effect for full body dyads that were configured to imply interactive behaviour, but not for dyads that were not. Specifically, subjects viewed pairs of briefly presented (30ms) human bodies or control objects (i.e. chairs), that either faced towards or away from each other, in either upright or inverted orientation, and were instructed to respond to which stimulus category they saw (i.e. bodies or chairs). Subjects were more accurate at detecting upright than inverted dyads when an interaction was implied by the two bodies facing towards each other, but not when facing away from each other. This effect was not shown with chair dyads, and indirectly suggests that upright interacting individuals may be processed in a configural, 'unified' manner, rather than as two separate individuals.

Similar interaction-based modulation of social perception has also been observed with face stimuli; emotional judgements of a target face are mediated by the interaction context implied by the relative orientation to another (non-target) face. Gray et al. (2017) used a psychophysical face-morph paradigm in which either a neutral, aggressive, or happy non-target face was presented next to a target face that varied on a fearful-happy expression continuum (e.g. 10% fearful, 20% fearful … 100% fearful), on each trial.

Crucially, interaction context was manipulated by whether or not the two lateral-view faces were configured to face towards or away from each other. The presence of a happy non-target face was found to bias judgements of the target face in the facing-towards, but not facing away condition. Together, these two studies demonstrate how implied interactions – through the fixed spatial relations of two bodies or faces – evoke perceptual biases that suggest a degree of shared, holistic, processing of interacting individuals, but not for non-interacting individuals. However, the influence of systematic differences in the spacing of critical body features is unknown (e.g. eyes or arms of the interactors are likely to be closer to each other when facing towards each other, than when not) and may at least partly contribute to these effects.

In addition to static stimuli, similar results are shown with dynamic interaction stimuli. Ding, Gao, and Shen (2017) found evidence for the 'chunking together' of interacting individuals' actions in working memory, but only for upright stimuli. On each trial, subjects were presented with a target action (i.e. an individual figure performing an action) that had also appeared in a preceding array, either as: One of several independent (and non-interacting) individuals; one half of an interacting dyad; or one half of a non-interacting dyad (i.e. performing incongruent interactive actions). Detection sensitivity to probed target actions was equivalent when the preceding memory arrays were comprised of either two separate individuals, or *two sets of interacting dyads*, but was significantly poorer for two non-interacting dyads (and for four individuals). Interestingly, this chunking effect was abolished in a follow-up experiment in which all of the same videos were *inverted* (i.e. sensitivity was greater for probed actions belonging to arrays containing two inverted individual figures than all arrays containing four figures – including those with two dyads). These memory effects complement perceptual inversion effects observed with static stimuli (Papeo et al., 2017).

The preceding findings demonstrate the importance of spatial configuration of dyads in perceiving interactions – specifically, facing direction and orientation. In addition to these 'fixed' cues, dynamic cues also aid in the detection and perception of interactive information. Neri, Luu, and Levi (2006) showed that observers were better at detecting the presence of human point-light dyads engaged in synchronised interactions than non-synchronised interactions; that is, detection was more robust when random noise dots were overlaid on intact dyads (engaging in fighting or dancing behaviour), than similar videos in which one figure's actions were time shifted (i.e. the two figures'

movements were desynchronised). Similarly, Thurman and Lu (2014) demonstrated the importance of motion congruency between agents when perceiving interactions. Specifically, above-chance accuracy for identifying interactions (from non-interactions) was shown for pairs of point-light figures when body form was scrambled, but congruency between the two figures' intrinsic and extrinsic movements (i.e. limb and average body movements, respectively) was preserved.

The importance of dyadic temporal correlations in perceiving interactive behaviour has also been shown with non-human moving geometric shapes (e.g. Heider & Simmel, 1944). For example, modulating the direction of movement by which one shape apparently pursues another, changes the degree to which the shapes' movements are perceived as interactive (Gao & Scholl, 2009). However, simple 'heat-seeking' chasing or purely correlated motion is not sufficient to induce this effect. Instead, it is evoked by dynamic changes in spatio-temporal contingencies between shapes that that imply animate, intentional behaviour (Bassili, 1976), that creates the 'automatic' perception of interactive behaviour (Scholl & Tremoulet, 2000). For example, changes in constant speed (Szego & Rutherford, 2007), average proximity (Roux, Passerieux, & Ramus, 2012), apparent facing direction (Gao, McCarthy, & Scholl, 2010), and small delays between shapes' respective movements (Schultz, Friston, O'Doherty, Wolpert, & Frith, 2005) all contribute to the perception of animate, interactive behaviour between shapes.

Across these studies, it is clear that the mere presence of multiple humans or agents does not constitute a meaningful interaction. Instead, specific configurations of spatio-temporal relations, such as facing direction and temporal correlations between interactors' movements, are crucial for the detection of interactive behaviour. Next, previous evidence that addresses the neural underpinnings of interaction perception is presented, specifically from functional magnetic resonance imaging (fMRI) studies.

1.5. Functional Neuroimaging of Social Interaction Perception

Compared to the vast amount of 'individual' person perception fMRI research, there is relatively little existent research that has investigated visual responses to social interactions. In keeping with the purpose of this thesis, this section focuses exclusively on studies that measured responses to third-person social interactions, rather than responses to first-person or second-person scenarios (e.g. Redcay et al., 2010) or gestures

performed by an individual person that could be construed as 'interactive' (e.g. Saggar, Shelly, Lepage, Hoeft, & Reiss, 2014). The first sub-section addresses studies that follow two main themes: 1) Studies that *quantitatively* vary the degree of interactive information, for example, comparing 'intact' interactive scenarios with similar control stimuli in which interactive cues have been disrupted, that constitute 'less-interactive' scenarios (e.g. interactors facing towards each other > facing away from each other); 2) Studies that compare *qualitatively* different 'types' of interaction, for example, affective vs. cooperative interactions. These two characteristics are directly related to the central questions of this thesis, however, an additional brief overview of studies that measure responses to other aspects of observed third-person interactions are presented (e.g. the incongruency of two interactors' actions). Additionally, the sample-size for each key study is reported for comparative purposes.

1.5.1. fMRI Research: Quantitative and Qualitative Interaction Comparisons

Responses to social interactions relative to single person scenarios elicit substantial activation differences across the brain. Iacoboni et al. (2004) showed subjects (N=13) 8*s* videos of two humans interacting, compared to 12*s* segments in which an individual was presented in isolation. This contrast (interactions > individual) yielded diffuse brain activation with notable peaks in middle temporal motion-sensitive area (MT), STS, precuneus, and dorsomedial prefrontal cortex (dmPFC). Similarly, Dolcos, Sung, Argo, Flor-Henry, and Dolcos (2012) presented subjects (N=18) with brief clips of animated interacting avatars and non-interacting avatars (i.e. two individuals approaching and greeting each other, or an individual approaching a cardboard cut-out figure of a human) and asked them to rate the avatars on several social dimensions (e.g. trustworthiness). Interactions evoked greater activation than non-interactions in the right pSTS and surrounding MTG/ITG, right IFG, amygdala, and dmPFC.

Both Iacoboni et al. (2004) and Dolcos et al. (2012) implicate the STS and dmPFC as regions centrally involved in processing social interactions. However, several limitations to these studies prevent such a clear interpretation. Firstly, these contrasts likely captured stimulus differences that were independent of interactive information; for example, the amount of body, face, and biological motion content all differed between conditions in both studies. This is particularly relevant given that regions in the STS are

sensitive to these types of information (e.g. Deen et al., 2015), and as such may confound this apparently 'interaction sensitive' response. Similarly, both studies showed strong responses in dmPFC – a key region in the mentalizing network – and may have captured higher level social cognitive processes that often covary with – but are not inherent to – interactive scenarios (i.e. unconstrained viewing of longer video segments in which top-down cognition typically arises, or when making explicit trait judgements about the interactors).

It is also worth noting that no fixed definition of 'non-interaction' was employed in these studies, highlighting that different interaction cues were disrupted to generate these control scenarios (e.g. Iacoboni et al., 2004 presented a single individual by themselves and thereby extinguished several sources of interactive information – for example, communicative intentions – that might have been present in the non-interaction stimuli used by Dolcos et al., 2012). Although the findings of Dolcos et al. (2012) may potentially provide a more precise measure of neural responses to social interactions, the intended purpose of this study was to measure responses to greeting behaviours between individuals, rather than to test responses to social interactions per se; it is also hard to know whether responses to this very specific 'greeting' scenario apply more generally to other interaction scenarios.

By contrast, other attempts have been made to directly compare responses to interactions and non-interactions, using more *visually constrained* stimuli. Centelles, Assaiante, Nazarian, Anton, and Schmitz (2011) used stimuli that manipulated facing direction along with the type of actions performed by individuals. Specifically, subjects (N=14) viewed displays containing two point-light human figures that either faced towards each other and interacted or faced away and performed separate non-interactive gestures. Interactions evoked stronger responses in the dmPFC, STS, and IFG than non-interactions. Although these stimuli were relatively 'visually matched', as with the preceding studies, these responses also likely include additional task-related processing (i.e. subjects made explicit top-down judgements about whether or not the two figures were interacting).

An even more constrained comparison of dyadic interactive responses has been demonstrated by Georgescu et al. (2014). In this study, action contingency was manipulated, such that two human-mannequin figures' movements were either: Contingent upon each other, as in a natural interaction; or 'mirrored', such that the two

figures movements were identical (i.e. the movements of figure A were duplicated and performed simultaneously by figure B). Subjects (N=28) showed strongest responses in the right posterior STG and right IFG (along with left parietal regions and fusiform gyrus) in the contingent interaction > mirrored interaction contrast. Unlike the preceding studies, no dmPFC activity was observed – potentially because the response task did not require social cognitive or person-directed judgements (i.e. subjects rated how 'natural' the scenes appeared to be).

Across these studies, the STS is routinely activated by human social interactions, but there is also evidence for similar sensitivity to non-human interactions. Castelli et al. (2000) used positron emission tomography (PET) to scan subjects (N=6) while viewing animations that depicted two moving shapes that either interacted together or moved randomly in an 'aimless' manner. A network of regions was found to respond more to interacting shapes compared to when the shapes moved randomly, with peak activity in the right pSTS, along with left STS, bilateral ITG, occipital cortex, and dmPFC.

This network of regions has been activated across a number of fMRI studies that used the same (Das, Lagopoulos, Coulston, Henderson, & Malhi, 2011; Gobbini et al., 2007; Kana, Keller, Cherkassky, Minshew, & Just, 2009; Tavares, Lawrence, & Barnard, 2007), or similar stimuli (Osaka, Ikeda, & Osaka, 2012) in neuro-typical individuals, as well as noted differences in several non-neurotypical populations such as autism (Kana et al., 2008) and schizophrenia (Das et al., 2011). However, these responses undoubtedly include explicit mentalizing judgements (e.g. 'how intentional do the shapes appear to be?'), as the focus of these studies has been to study mental state attribution rather than interaction perception. The implications of using explicit judgement tasks in conjunction with these abstract moving shape interactions is particularly relevant to regions such as the pSTS. Lee et al. (2012) demonstrated strong right pSTS activity when participants made explicit intentional attributions (i.e. detecting which shape chased another) for randomly moving shapes. In comparison, this was not the case when subjects made non-social judgements about the shapes movement (e.g. speed).

In comparison to the preceding studies that have made quantitative comparisons between interactions and 'less-interactive' control stimuli, Sinke, Sorger, Goebel, and de Gelder (2010) compared responses to *qualitatively* different interactions. Subjects (N=14) viewed brief videos that depicted either a threatening or teasing interaction (i.e. a male actor either aggressively or playfully tried to snatch a handbag from a female

actor), and performed either: An explicit judgement task or orthogonal response task (i.e. identifying the emotional tone of the video, or judging the color of dots superimposed on the video, respectively). Greater responses for the threatening > teasing contrast were observed in right STS, left FG, right EBA, and left IFG but only when performing the orthogonal response task, demonstrating that this differentiation is unlikely driven by explicit higher cognitive social judgements. Additionally, similar responses were found in the right amygdala but for both the explicit judgement and orthogonal response tasks. These findings suggest that these regions may play a role in differentiating between qualitatively different interaction scenarios. However, it is also worth considering that similar responses have been observed with single body stimuli (Kret, Pichon, Grèzes, & de Gelder, 2011), suggesting that valence information (i.e. negative > positive emotional contents) that is not specific to social interactions might contribute to this effect.

In comparison to studies using dynamic interactions as stimuli, several studies have also used static image stimuli to probe social interaction responses. Kujala, Carlson, and Hari (2012a) showed subjects (N=19) images of two human individuals either facing towards or away from each other and instructed them to freely inspect the images. For the facing towards > facing away contrast, diffuse activation differences were observed with notable peaks in the bilateral amygdala, pSTS, and dmPFC, along with IPS, premotor cortex and supplementary motor area. Interestingly, highly similar cortical responses (i.e. similar cortical overlap) in dog experts (but not control participants) was also shown in response to images of two interacting dogs facing towards each other, than when facing away, suggesting the importance of visual experience in interaction perception (Kujala, Kujala, Carlson, & Hari, 2012b).

However, the effects in these two studies may be driven by other social interaction cues besides facing direction. For example, physical contact between interactors, or qualitative differences in poses that were not matched between stimuli (e.g. communicative gestures in the facing toward but not facing away stimuli). Indeed, Quadflieg, Gentile, and Rossion (2015) observed no response differences in any brain region with a similar facing towards > facing away contrast in which the distance between interactors was closely controlled, and the same postures were presented in both conditions.

Similarly to Sinke et al. (2010), Canessa et al. (2012) compared responses to qualitatively different dyadic interactions, using photograph stimuli. Subjects (N=27)

viewed photographs of affective and cooperative interactions (i.e. two individuals demonstrating positive affectionate behaviour, or helping each other perform a task together, respectively) and were instructed to give a button-press response whenever a control image of a landscape was presented. Conjunction analysis revealed common activation in the right TPJ and superior orbital gyrus, for both interaction types. For the cooperative > affective contrast, diffuse activation was observed with peak activation in occipitotemporal areas, whereas the reverse contrast revealed activation in the mPFC only. It is worth noting that these stimuli differed on several other 'non-interactive dimensions', that may account for these differential responses. For example, cooperative stimuli contained relatively more object-action information (e.g. two individuals lifting a bucket of water together) that is consistent with the sensitivity of the LOTC to static images of different human action information (Hafri et al., 2017). Therefore, it is hard to know to what extent these responses reflect sensitivity to different types of interactive information per se, or whether they are driven by these stimulus differences.

To summarize, across most of these dyadic third-person interaction studies, the STS is routinely activated. Similarly, the IFG and dmPFC are also frequently recruited. However, there are several crucial methodological issues across these studies that prevent a clear and unequivocal account of the role of these regions in *visually perceiving* social interactions. For example, many of these studies varied considerably in terms of how stimuli were manipulated (i.e. which interactive cues were altered), analysis sensitivity (e.g. relatively weaker sensitivity for whole brain vs. region of interest (ROI) analysis) and sample size (e.g. several studies had $N < 20$ subjects).

Additionally, as already stated, most of these studies did not use an orthogonal response task that may have served to limit the contribution of top-down social cognition that might have driven responses in dmPFC. Indeed, studies that probe higher level cognitive responses to third-person social interactions in the *absence of perceptual person information* report stronger dmPFC responses (and also in other mentalizing network regions) when simply imagining interactive behaviour (Trapp et al., 2014), or when reading written stories that cue mentalizing (Walter et al., 2004) relative to non-interactive comparison conditions.

This suggests that dmPFC responses are not centrally involved in the *visual perception* of social interactions. However, that is not to say that dmPFC does not play a key role in more naturalistic observation of interactions, as underscored by strong

correlations between activation and the presence of social interactions during natural viewing of extended movie sequences (Wagner, Kelley, Haxby, & Heatherton, 2016). Similarly, recent evidence demonstrates the central importance of medial and ventro-lateral frontal regions (as well as temporal and parietal networks) in response to conspecific interactions viewed by rhesus macaque monkeys (Sliwa & Freiwald, 2017).

1.5.2. fMRI Research: Additional Interaction Comparisons

One important aspect of understanding and recognising interactions is the perceived congruency between interactors. Quadflieg et al. (2015) investigated the role of congruency between interactors respective actions. Subjects (N=12) performed a sex categorisation task while viewing images of human dyads (i.e. subjects were asked to determine whether each dyad was comprised of two individuals of the same or different sex). Three different types of dyads were presented: Congruent and incongruent dyads (i.e. performing actions complementary to a given interaction, or paired individuals from separate, non-complementary interactions, respectively), along with non-interactive dyads where two individuals from each congruent interaction were repositioned to *face away* from each other. No differences were observed for the congruent > non-interaction contrast in any brain region. Instead, for the incongruent > non-interaction contrast, bilateral MTG, right parahippocampal gyrus and left pSTS was found, and similar differences were observed when contrasting incongruent with congruent interactions in several temporal lobe regions (i.e. bilateral FFA, FBA, EBA, and pSTS).

These findings broadly demonstrate sensitivity to dyadic (in)congruency information in several posterior temporal lobe regions, and may reflect 'additional visual processing' required to resolve incongruent dyadic information, similar to previously observed social visual brain responses to incongruent stimuli (e.g. Vander Wyk, Voos, & Pelphrey, 2012). By contrast, incongruently paired dynamic interactors (i.e. point-light dyads performing actions taken from different interactions) evoke strong responses in the precuneus and several frontoparietal regions relative to congruent interactors (Petrini, Piwek, Crabbe, Pollick, & Garrod, 2014) – responses that the authors attribute to violations in social expectations of interactors. These findings do not conceptually align with evidence in the previous section that generally make 'intact' > 'disrupted' interaction comparisons. Instead, incongruent > congruent contrasts could be construed as

capturing the opposite comparison (i.e. 'disrupted' incongruent > 'intact' congruent interactions). Therefore, these findings might suggest which regions play an important role in resolving ambiguous interaction scenarios.

Another interesting aspect of social interactions concerns how the perceived agency of interactors – for example, whether they are human or robot – alters how they are perceived or understood. Wang and Quadflieg (2015) showed subjects (N=26) images of humans either interacting with another human or a robot and instructed them to judge whether or not one of the interactors was helping the other. ROI analyses showed greater response to human-human than human-robot interactions in the right FFA, bilateral pSTS, and left TPJ. Greater responses to human interactions in these regions are partly compatible with previous responses to conspecific interactions, that are likely the result of greater visual experience than with non-conspecific interactions (Kujala et al., 2012b). By contrast, greater responses in medial mentalizing regions – the precuneus and ventro-medial PFC – for the human-robot > human-human contrast, may reflect increased top-down processing of these unusual scenarios.

Finally, in contrast to using dyadic interaction stimuli, Huis in 't Veld and de Gelder (2015) investigated responses to interactive compared to individual behaviour within the context of a crowd. Specifically, subjects (N=16) viewed videos in which crowd members expressed either happy, fearful, or neutral behaviours by themselves, or with others (i.e. individually or interactively), while performing an orthogonal colour judgement task (i.e. make same/different judgements for two dots briefly overlaid on videos). The interactive > individual contrast revealed widespread whole brain activation with notable peaks in several bilateral occipitotemporal regions, right middle temporal gyrus, precuneus, pre/postcentral gyri, and left STS. Although interactive and individual conditions differed across a number of dimensions (e.g. greater stimulus motion energy as well as greater behavioural ratings for arousal and valence for interactive conditions) – the authors assert that these findings demonstrate how complex interactive scenes evoke enhanced responses in numerous AON regions that are typically also recruited when viewing actions performed by individuals.

Together, these diverse findings demonstrate the nuanced activations in social perception and cognition network regions that have typically been associated with individual agent processing, and demonstrate the scope for further research to explore fMRI responses to social interactions.

1.6. Developmental Social Visual Perception

A very brief overview of developmental social vision research is presented here, to provide context for the experiment presented in chapter 4.

1.6.1. Static Social Visual Perception

Developmental behavioural studies of face and body perception show that young children demonstrate relatively similar responses to those shown by adults. For example, 'infant looking paradigms' (i.e. where the length of time that an infant spends looking at a target stimulus relative to a simultaneously presented non-target stimulus serves as a proxy measure for recognition, or some other perceptual ability) provide evidence of configural face processing. For example, new-born infants (< 8 days) stare for longer at attractive faces than less-attractive faces (as determined by adult attractiveness ratings; Slater, Quinn, Hayes, & Brown, 2000). Similarly, 3 – 4 month-old infants stare for longer at female than male faces (Quinn, Yahr, Kuhn, Slater, & Pascalis, 2002). Crucially, however, these effects are abolished when faces are inverted, and therefore configural processing is disrupted. Configural processing of bodies is also observed in infants (< 4 months old), as shown by longer staring times for normally configured bodies than bodies with configural distortions (e.g. arms positioned on the legs rather than the torso, or irregular scaling of the torso; Zieber, Kangas, Hock, & Bhatt, 2014).

Conversely, children also show relatively immature social visual responses, relative to adults. For example, sensitivity to the changes in the spacing of internal facial features is not 'adult-like' by 8 years of age (Mondloch, Dobson, Parsons, & Maurer, 2004), and event-related potential (ERP) correlates of emotional face processing are not fully mature until late adolescence (Batty & Taylor, 2006). Indeed, developmental fMRI studies typically show nuanced differences in functional responses to face and body information between children and adults. For example, cortical responses to faces in 4 – 6 month-old infants are localized to similar regions as adults (e.g. FFA, OFA, STS), however the magnitude of these responses are substantially smaller, and these regions do not exhibit the same strong selectivity for faces (i.e. face responses are not greater than for other categories, such as bodies, objects, and scenes) that is seen in adults (Deen et al., 2017). Similarly, increases in the selectivity and size of FFA are observed with increasing age

(between 7 – 32 years), yet these cortical changes are *functionally specific* to face responses and such trends are absent in overlapping body selective FBA cortex (Peelen, Glaser, Vuilleumier, & Eliez, 2009). As for body perception, greater responses are reported in adults than children, in FBA and other body selective regions (e.g. EBA and pSTS; Ross, de Gelder, Crabbe, & Grosbras, 2014). These findings suggest that some person perceptual abilities are present in early infancy, while others undergo further development, as supported by developmental differences in functional cortical responses.

1.6.2. Dynamic Social Visual Perception

As with static face and body processing, young infants also show sensitivity to the processing of relatively more complex, abstract human point-light biological motion stimuli. For example, infants (< 5 months) stare for longer at intact point-light human figures compared to spatially scrambled versions of these figures when they are presented in upright, but not inverted orientation (Bertenthal, Proffitt, & Kramer, 1987). Similarly, early ERP responses are highly similar to adults in 8 month-old infants (i.e. differences between intact and scrambled point light stimuli occur 200 – 400ms after stimulus onset; Hirai & Hiraki, 2005). These findings are somewhat surprising, given the complex nature of these stimuli (i.e. body structure is 'constructed' from motion information, rather than from static body information).

Sensitivity to biological motion is also a pre-requisite for perceiving dynamic social interactions, and there is evidence for developmental changes in perception of interacting human point-light dyads. Centelles, Assaiante, Etchegoyhen, Bouvard, and Schmitz (2013) showed typically developing children (N=36) and adults (N=12) videos (as used by Centelles et al., 2011) of interacting point light dyads and non-interacting dyads (i.e. that performed separate, non-interactive, and non-communicative gestures), and instructed them to make button-press judgements as to whether or not the figures in each dyad were communicating with each other. Younger children (4 – 6 years) were significantly less accurate at performing this task overall and were also less accurate for interaction than non-interaction (this was also true for children aged 7 – 8 years). These findings suggest suboptimal perception and categorization of interactive behaviour in younger children, relative to older children (9 – 10 years) and adults.

fMRI responses to these stimuli were also compared between children and adults in a recent study. Sapey-Triomphe et al. (2017) presented these stimuli to children (8-11 years), adolescents (13-17 years), and adults (20+ years), while they judged whether the two figures were acting together or separately. For the interaction > non-interaction contrast, all 3 groups showed similar activations in pSTS, anterior and posterior MTG, middle occipital gyri (MOG), inferior temporal gyrus (ITG) and IFG. Main effects of age-group showed diffuse brain differences, that included peaks in the bilateral MTG, STG/STS, and IFG, as well as right lingual gyrus. The authors also analysed parametric changes in activation for this contrast with increasing age, and found greater fronto-parietal activity (e.g. middle frontal, precuneus, and IPL) with increasing age, and greater temporo-occipital activation with decreasing age (e.g. ITG, MTG, and STS). This broad distinction in regional activation across age aligns with the possibility that adults recruit more mentalizing and attentional resources (although this may have been driven more by the social judgement task than interaction perception per se), whereas children are more reliant on temporo-occipital regions while observing social interactions.

In summary, young children show a remarkable ability to perceive and understand the presence of visually constrained biological motion displays, and are able to extract interactive information from dyadic point-light stimuli. However, these perceptual abilities – and their subserving neural responses – appear to undergo considerable maturation across childhood and adolescence and are deserving of further investigation.

1.7. Thesis Overview

1.7.1. Motivation for the Current Research

The empirical research undertaken in this thesis aims to characterize brain responses to visually observed social interactions. These experiments are motivated by several crucial insights presented over the course of this chapter. Firstly, that behavioural evidence of configural inversion effects, 'chunking together' of interactors' actions, and the manipulability of apparent interaction cues (e.g. facing direction and action contingency) demonstrate that interactions are not merely perceived as multiple individuals, but instead as 'interacting units'. Secondly, few fMRI studies have specifically

sought to isolate visual responses to social interactions, and most contain potentially confounding, extraneous stimulus and task-related processes. Thirdly, few of the existing social interaction studies have employed highly sensitive analysis procedures (e.g. the use of functionally defined ROIs, or multivoxel pattern analysis (MVPA) approaches) to allow for the detection of highly nuanced social interaction responses. Therefore, a principled investigation of the neural responses to visually observed third-person social interactions is needed.

1.7.2. Overview of Chapters

Chapter 2 is a general methods section in which an overview of basic fMRI acquisition and analysis is described, along with justification of the specific fMRI designs and analyses used in this thesis.

Chapter 3 is an empirical fMRI chapter in which the following research questions are asked:

a. Which brain regions play a central role in perceiving third-person social interactions?
b. Which regions are more responsive to social interactions than non-interaction control stimuli?
c. Are such regions also sensitive to the contents of interactions?
d. Are these regions sensitive to non-human interactions?

Chapter 4 is an empirical fMRI chapter in which the following research questions are asked:

a. Do children show weaker 'interaction selective' responses than adults?
b. Are interaction selective regions less sensitive to other social information, such as faces and bodies – in both adults and children?

Chapter 5 is an empirical fMRI chapter in which the following research questions are asked:

a. Are there brain regions that represent interacting dyads represented as 'more than the sum of their parts'?

b. Which type(s) of social information might these regions integrate?

Chapter 6 is a general discussion in which the key findings across the three empirical chapters are synthesized and interpreted.

Chapter 2

General Methods

2.1. Overview of fMRI Methods

2.1.1. Basis of the BOLD Signal

Functional magnetic resonance imaging (fMRI) studies were first conducted in the early 1990s (Bandettini, Wong, Hinks, Tikofsky, & Hyde, 1992; Kwong et al., 1992; Ogawa et al., 1992) as a means of measuring changes in brain activity in response to a controlled stimulus (e.g. flashing visual stimulus). Inference of brain activity relies on task-related changes in the blood-oxygen-level-dependent (BOLD) signal (Logothetis & Wandell, 2004); briefly explained, *relative increases* in the ratio of *paramagnetic* deoxygenated haemoglobin and *diamagnetic* oxygenated haemoglobin that arise from increased metabolic processes associated with neural activity, result in detectable signal changes. This task-related change in MRI signal is interpreted as increased local activation. Although BOLD signal is an indirect measure of brain activation (i.e. it is correlated with several neural processes, such as action potentials and post-synaptic activity; Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001), it is considered to be a reliable approximation of increased local neuronal activity (Huettel, Song, & McCarthy, 2004; Logothetis et al., 2001; Soares et al., 2016).

The BOLD signal is characterized by haemodynamic response function (HRF) which typically corresponds to a 5 – 6 second delay in signal increase after stimulus onset, that returns to baseline around 12 seconds after stimulation, followed by a small negative-going decrease before returning to baseline (Soares et al., 2016). This 'canonical' HRF signal is typically convolved with a time series corresponding to stimulus onsets across a scan, to estimate the degree to which BOLD signal changes in a given brain area are modulated by stimulus presentations. Despite some inter-subject variability in the shape of HRF signal (Handwerker, Ollinger, & D'Esposito, 2004), the canonical HRF response modelled in typical neuroimaging software packages such as SPM (fil.ion.ucl.ac.uk/spm/software/spm12) provide a relatively robust estimate of BOLD signal (e.g. Zumer, Brookes, Stevenson, Francis, & Morris, 2010).

During typical fMRI acquisition sequences, BOLD signal is sampled across the brain (and later reconstructed) as a set of contiguous cubic volumetric elements called 'voxels'. As such, a raw fMRI brain volume is represented as a 3D matrix of voxels where the intensity value at each voxel corresponds to the BOLD signal measured at that

location (Huettel et al., 2004). Each volume is acquired as a sequence of 'slices' (e.g. 30 ascending horizontal slices). The acquisition time for a single volume varies with different acquisition parameters but is generally in the order of 2000ms. Therefore, a given acquisition run is comprised of a series of brain volumes. Despite this relatively poor temporal resolution, considerably robust estimation of the haemodynamic response is possible, allowing for reliable inference of functional brain activation in many scenarios.

2.1.2. fMRI Pre-processing and Univariate Inference

Before fMRI data is analysed, a given subject's 'raw' time series data (i.e. run-wise sequences of whole brain BOLD images) are typically subjected to several pre-processing steps as follows (Huettel et al., 2004; Strother, 2006): Slice-timing correction (i.e. using temporal interpolation to minimize timing differences across sequentially acquired slices); rigid-body realignment (i.e. correcting head motion across volumes by realigning volumes across three translation and three rotation parameters); co-registration (i.e. co-registration or realigning the subject's anatomical image with their corresponding fMRI time series data); segmentation (i.e. generating separate grey-matter, white-matter and other tissue maps that are later used for image normalization); normalization (i.e. non-linear warping of images to a common template brain); smoothing (i.e. low-pass spatial filtering of images to boost signal-to-noise ratio); high-pass temporal filtering (i.e. removing low-frequencies that reduce signal to noise ratio, such scanner-related and physiological signal drift). These steps are not mandatory for all analyses, but are generally necessary for most univariate fMRI analyses.

General linear modelling (GLM) is performed on pre-processed time series data – that is, images that have been realigned, normalized, and smoothed (Soares et al., 2016). Specifically, GLM is performed at each voxel, by estimating the regression coefficient from the comparison between 'actual' BOLD signal and predicted BOLD changes (i.e. HRF-convolved stimulus onset time series); this results in the creation of a series of beta maps where each voxel represents a beta coefficient. When estimating responses to several experimental conditions (e.g. social interactions and non-interactions), one beta map per condition is created for each run, along with an intercept image and further maps for each of the six head motion parameters.

To make univariate comparisons between conditions, beta maps are used to create further contrast images; for example, for a contrast between interaction and non-interaction conditions, a contrast image is created where voxel-wise subtraction of run-wise non-interaction images from interaction condition is performed. In SPM, a corresponding image of t-values (i.e. t-map, or f-map in the case of an ANOVA) is also generated along with a contrast image and is then subject to inferential testing to determine which brain regions are activated or demonstrate measurable sensitivity to a given contrast. It is worth noting that maps for other statistics (e.g. Pearson's *r* or pattern classification accuracy (%) etc.) can also be created in similar ways and used to make similar inferences. Such inferences can be made at the subject level, or combined at the group level, by aligning subjects' normalized images (i.e. statistical maps that have been registered to a common brain template).

One important consideration when making inferences on whole brain data is the multiple comparison problem; due to the large number of voxels (e.g. typically between 10'000 to 20'000 voxels; Forman et al., 1995; Huettel et al., 2004) and the large number of associated voxel-wise inferential tests (e.g. t-tests), family-wise error (FWE) correction of false-positive activation results in extremely conservative activation thresholds (e.g. .05 / 20,000 voxels = .0000025). As such, only voxels with very strong activation (and correspondingly very low *p*-values) 'survive' correction. Instead, more lenient false-discovery rate (FDR) correction (Benjamini & Hochberg, 1995) is possible, whereby a similar correction procedure to FWE is applied, but instead of considering all brain voxels, a correction criterion is applied only to 'significant' voxels (e.g. all those with an uncorrected *p*-value < .05), therefore allowing for considerably more liberal correction in many scenarios.

Although FDR correction affords relatively more lenient thresholding than FWE correction, and as such reduces the number of false-negatives, typical whole brain analyses can yield up to several thousand 'significant' voxels that still results in relatively conservative correction of activation. Instead of correcting on a voxel-wise basis, cluster-wise thresholding (Forman et al., 1995; Xiong, Gao, Lancaster, & Fox, 1995) in conjunction with principles of Gaussian random field theory (Worsley et al., 1996) offer a more lenient means of multiple comparison correction of whole brain activation. This approach is motivated by the spatially extended nature of BOLD activation – that is, that activation tends to form clusters of similarly activated voxels – and so cluster-wise inference is often

more appropriate than voxel-wise correction (Huettel et al., 2004). Therefore, the use of FDR cluster-wise correction allows for valid whole brain activation inference while reducing the number of false-negative activations.

An alternative (and complementary) approach to whole brain analysis is functional region of interest (ROI) analysis (e.g. Saxe, Brett, & Kanwisher, 2006). This broad approach entails functionally localizing and, subsequently testing responses within, regions of cortex that are hypothesized to be sensitive to a given task or stimulus condition. For example, to test responses to the different types of social interaction (e.g. competitive or cooperative interactions) in cortex theorized to be sensitive to social interaction information, an interaction 'localizer' task could be used by contrasting responses to interaction videos with non-interaction videos (e.g. social interactions > non-interactions). Once these regions are identified, a separate task (or separate data) can then be used to test responses in this region (e.g. comparing videos of cooperative and competitive interactions). This approach carries two main benefits over whole brain analyses (Saxe et al., 2006): Firstly, substantially greater statistical power is afforded than with whole brain analysis as highly conservative multiple comparison correction is not required; and secondly, this approach allows for inter-subject variability in the location of activation (or ROIs) without obscuring group level activation trends; by comparison, such variability may result in poor estimation in whole brain analyses, due to weak overlap in activation across subjects. Additionally, a single statistic can be calculated across all voxels within and ROI (e.g. mean percent signal change) which can then be easily compared with some other measure of interest on a subject-wise basis (e.g. a correlation between subjects' mean neural response and a behavioural rating score).

However, restricting analyses to a-priori defined ROIs may also result in overlooking other brain regions that may be sensitive to a given manipulation. Therefore, in many cases it is ideal to employ both ROI and whole brain analyses. Further to this, one crucial consideration for valid inference from ROI analysis is that the data used to define an ROI must be independent of the data used to extract responses (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009). Failure to ensure this independence (also known as 'double dipping') produces spurious or circular results, as responses are biased by the ROI voxel selection criteria, and vice versa. To overcome this, independent data from either a different task, different subject, or different run of the same task, must be used for ROI definition and extraction.

2.2. fMRI Methodology in This Thesis

2.2.1. Justification for the Use of fMRI

The use of fMRI in this thesis is justified by the overarching aim of the presented experiments. That is, to determine which brain regions are strongly responsive to – and therefore play an important role in – social interaction perception. fMRI is most suitable to address this aim, due to the relatively high spatial resolution it affords compared to other approaches, such as electroencephalography (EEG). However, EEG, in conjunction with fMRI, may serve to further characterize the neural processes that underlie social interaction perception (e.g. the high temporal resolution of EEG approaches may contribute to understanding the timescale of such responses). Indeed, the use of fMRI to identify and characterize key regions that are sensitive to social interactions may aid in the interpretability of results from other approaches with poorer spatial resolution.

2.2.2. fMRI Designs

Two different fMRI designs were employed in this thesis. For chapters 3 and 4 (and for localizer tasks in chapter 5) a blocked design was used – this was motivated by the strong detectability of BOLD signal that these designs afford. Specifically, blocked designs exploit the superposition of haemodynamic responses in 'blocked' sequences of individual trials (i.e. the responses of multiple 'temporally proximal' stimuli are linearly summed together). This typically results in strong, sustained activation for the period of stimulation in regions that are sensitive to the information in question (Huettel et al., 2004). By contrast, responses are not 'summed together' and hence are considerably weaker for brief, individually presented stimuli. Blocked designs therefore produce strong, detectable activation changes to a given stimulus condition. It is suggested that block lengths greater than 10s along with 'rest' blocks (i.e. no stimulation) of equal duration are necessary for robust detection of BOLD signal changes (Huettel et al., 2004).

By contrast, an event-related design was used for the main task in chapter 5. The motivation for this was to allow for the analysis of responses to individual stimuli (as well as stimuli grouped together as conditions). In contrast to strong detection of activation in blocked designs, event-related designs afford greater estimation of the timing and shape

of haemodynamic responses to individual stimuli (Huettel et al., 2004). Due to the slowness of the haemodynamic response, designs with fixed interstimulus intervals must have brief trials spaced approximately 10 – 12s apart in order to accurately estimate stimulus-wise haemodynamic responses (i.e. to allow sufficient time for the haemodynamic response baseline stabilization before the onset of the next trial; Bandettini & Cox, 2000). However, estimation is improved by systematically randomizing the event design; as such, 'jittering' the interval length between stimuli and randomizing the order of stimulus presentation allows for stimulus signal sampling at *different timepoints of the haemodynamic response* across an acquisition run, resulting in better estimation of individual, yet overlapping haemodynamic responses to stimuli (Burock, Buckner, Woldorff, Rosen, & Dale, 1998; Miezin, Maccotta, Ollinger, Petersen, & Buckner, 2000).

2.2.3. Univariate ROI Analyses

In accordance with the benefits of functional ROI analysis (Saxe et al., 2006), the experiments in this thesis focused predominantly on this approach. For univariate ROI analyses in chapters 3 and 4, percent signal change (PSC) was used to measure ROI responses to different conditions. This approach was chosen over simple comparison of beta weights between conditions (i.e. comparing beta weights for different conditions within an ROI, or across ROIs) for the following reason. Beta values (i.e. voxel values within a given beta map) are generated during GLM estimation based on *global* mean scaling of the whole brain, but because BOLD responses are variable across different tissue types (e.g. greater responses in grey matter than white matter and cerebrospinal fluid), such *global* estimates of signal change are not best-suited for calculating *local* percent signal change in ROIs (Gläscher, 2009). This is especially true when comparing responses between different regions of the brain that might vary in their baseline activation (or the amount of grey matter in these regions). Instead, ROI responses to stimulation blocks (or individual events) should be based on estimates of baseline activation *within* a given ROI. In accordance with this logic, univariate PSC for ROIs in this thesis are based on the following equation, as implemented with the MarsBaR toolbox (Brett, Anton, Valabregue, & Poline, 2002):

$$\frac{(\beta\ \text{task} * \max(\text{HRF})) * 100}{\beta\ \text{intercept}}$$

Specifically, to calculate the PSC for a given activation block, at a given voxel, the corresponding beta is multiplied by the maximum height of the canonical HRF (as estimated by the duration of stimulation). The resulting value is then multiplied by 100 and then divided by the mean response value of a given voxel (i.e. the corresponding value from the intercept image that is generated during GLM estimation). As such, the resulting PSC value represents the change in voxel signal for a given block, relative to the mean voxel response across the entire run (i.e. across all conditions and rest). This value is generated across all voxels within an ROI, and averaged across all instances of a given condition across experimental runs as a measure of PSC.

2.2.4. Multivariate Analyses: A Brief Overview

In addition to univariate analyses, multivoxel pattern analyses (MVPA) were also performed. MVPA refers to a broad category of approaches whereby inference is based upon multivariate distributed voxel responses. The motivation for such analyses is demonstrated in the following example. A given ROI may be theorized to play an important functional role, such as recognizing gender from human body or face information. For a univariate comparison, the *mean* response across voxels can be extracted for both male and female stimuli and compared with each other (e.g. female > male; male > female), however, in such scenarios, it is not intuitively clear whether female or male stimuli should evoke greater activation. Indeed, such comparisons may show that there is no univariate difference between conditions. However, it is possible that the *patterns* of response across voxels are consistently differentiable between the two categories, and so in this case, comparisons of distributed voxel responses between conditions may offer greater sensitivity than univariate comparison (Norman, Polyn, Detre, & Haxby, 2006).

The first published use of MVPA used a correlation approach (Haxby et al., 2001). This entailed extracting ventral temporal lobe voxel response patterns for images of faces, houses, and other categories of objects and comparing within category (e.g. face to face) with between category (e.g. face to house) pattern correlations. Significantly higher

correlations for within category patterns than between category patterns was interpreted as 'above-chance' differentiation of object categories in this broad region of cortex. As such, the ventral temporal lobes contain distributed patterns of activation that reliably differ across object categories.

2.2.5. Multivariate Analyses: SVM Classification

The underlying logic of the approach used by Haxby et al. (2001) – that is, to test whether regions of cortex contain voxel patterns that significantly differentiate or *classify* between different categories of information – has been applied widely to many novel neuroimaging experiments, that use various different statistical approaches and algorithms (e.g. Misaki, Kim, Bandettini, & Kriegeskorte, 2010; Coutanche, Solomon, & Thompson-Schill, 2016). In the current thesis, linear support vector machine (SVM; e.g. Chang & Lin, 2011) classification was employed to address whether certain brain regions show evidence of categorical differentiation of responses to video stimuli. The current use of SVM classification was motivated by empirical evidence that it performs equivalently or better than many other classifiers (Misaki et al., 2010).

A brief overview of linear SVM classification is outlined, as used with fMRI data in the current thesis, with a two-category example of classification. For simplicity, this example considers a pattern of just two voxels, but patterns are typically in the order of > 50 voxels. For a given subject, patterns of data are extracted from the same set of voxels for class A and class B. A subset of data is then used to train the SVM classifier to 'learn' the associations between a given class label (e.g. social interaction) and the voxel patterns of data that correspond to it. Accordingly, patterns are mapped to high-dimensional space. Figure 2.1. shows each pattern (or *example*) as a single data point with 2 dimensions (i.e. *n* voxels = *n* dimensions); larger patterns are therefore harder to visualize, given the high number of dimensions. Classification is achieved by the calculation of an optimal decision boundary (or *hyperplane*), that is typically based on the maximum margin between the examples of each class that are closest to the other class. The examples used to determine the position of this boundary are known as 'support vectors'.

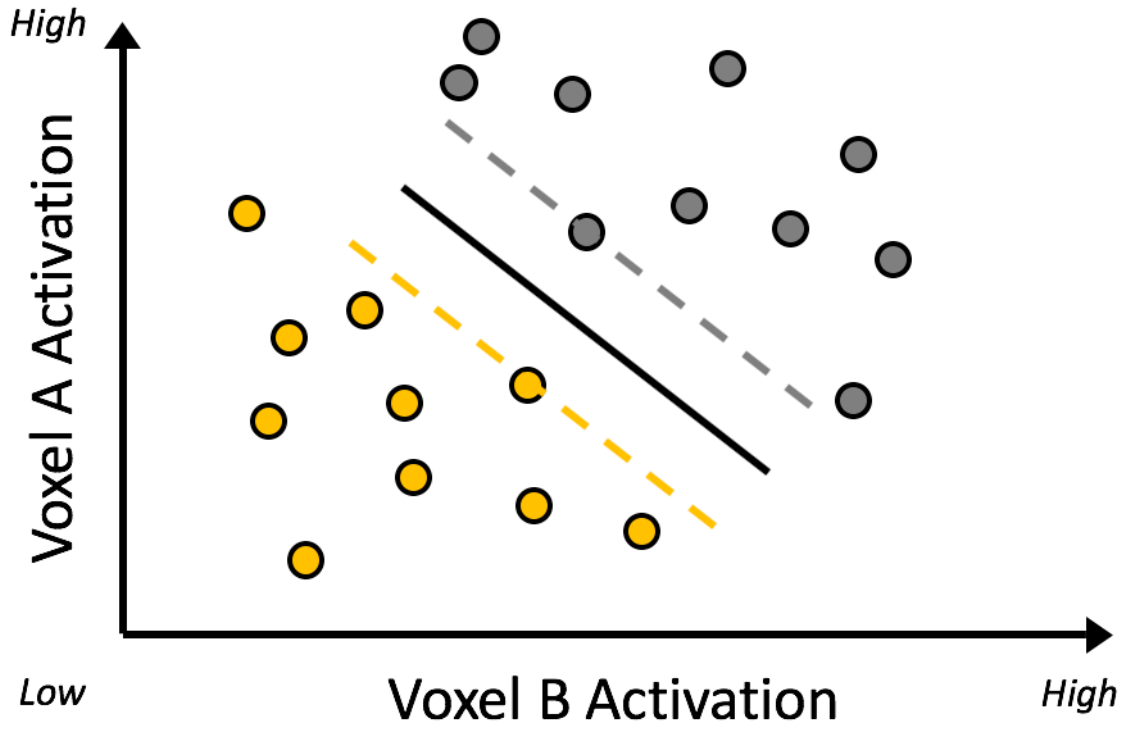


Figure 2.1. A simplified illustration of SVM classification between two classes (orange and grey circles) based on the activations of just two voxels; each example (coloured circle) corresponds to the activation values for voxel A and voxel B. The bold black line represents the hyperplane (decision boundary). Dashed-lines correspond to the maximum margins.

Generalisation of classification (*cross-validation*) is then performed whereby the 'trained' decision boundary is then applied to an independent set of data to test classification performance, often expressed as a percentage accuracy score:

$$\left(\frac{correctly\ predicted\ examples}{total\ tested\ examples}\right) * 100$$

This value therefore represents the percentage of independent voxel patterns that were correctly classified across all classes. However, 'class-wise' accuracies can also be obtained to determine whether accuracy differs between classes. In chapter 3, a simple 'one-against-one' approach was employed where a single classifier was trained to differentiate the two classes of interest. However, for chapter 5, classification of three classes was implemented with three separate 'one-against-one' classifiers for each unique pairing of the classes. In this case, classification accuracy was calculated as the

averaged accuracy for each class, across classifiers. It has been demonstrated that in many cases this approach performs better than a 'one-against-the-rest' procedure (e.g. one class against the remaining other classes, iterated across classes; Hsu & Lin, 2002).

SVM classification was performed both within ROIs and for whole brain analyses. For ROIs, this simply entailed performing classification on within-ROI voxels only, and reporting the mean classification accuracy across subjects, for each ROI. For whole brain analyses, a 'searchlight' procedure (Kriegeskorte, Goebel, & Bandettini, 2006) was employed. This entailed placing a 'searchlight' sphere (e.g. 5 voxel radius) centred on a given voxel and performing classification across all searchlight voxels. The resulting classification accuracy is then mapped to the corresponding central searchlight voxel on an output searchlight map. This procedure was iterated across every voxel until a whole brain searchlight map was generated. This was performed on a per subject basis, such that subject searchlight maps were then entered into a group level analysis for statistical inference.

2.2.6. Multivariate Analyses: Representational Similarity Analysis

In addition to SVM classification, another MVPA compatible approach – representational similarity analysis (RSA; Kriegeskorte, Mur, & Bandettini, 2008) – was used in chapter 5. Specifically, for voxel patterns of data, RSA entails the construction of (dis)similarity matrices in which each cell represents the similarity (or inversely, dissimilarity) of responses between each given pair of stimuli, conditions, or items. Typically, this is expressed as a series of pair-wise correlations between voxel patterns, that together express the representational structure of information in a given brain region. One main advantage of creating similarity matrices – and therefore expressing information in 'representational space' – is that comparisons are easily made with similarity matrices generated from *different voxels* in other brain regions; two matrices can then be vectorized and correlated with each other to determine the overall similarity in representational structure. Further to this, similarity matrices can also be constructed from univariate measures (e.g. mean PSC, or behavioural Likert ratings) by calculating the absolute difference between pairs of individual values (e.g. rating for stimulus A vs. rating for stimulus B). Therefore, RSA allows for the comparison of representational contents of otherwise incomparable data, based on separate brain regions or measures.

In chapter 5, RSA was used to determine the representational similarity between ROI voxel pattern responses to different but related sets of stimuli.

2.2.7. Whole Brain Inference

In addition to ROI analyses, univariate whole brain analyses were also conducted in chapters 3 and 4. For these analyses, a height threshold of $p = .001$ (i.e. only strongly activated voxels with corresponding p-values < .001 surviving this initial threshold were retained) and FDR cluster correction threshold of $p < .05$ was employed. The use of these thresholds is motivated by the comparability to other studies that used similar thresholds (Isik, Koldewyn, Beeler, & Kanwisher, 2017; Sapey-Triomphe et al., 2017). Further to this, previous evidence shows that these thresholds offer relatively good control of false-positive rates for blocked design data (Eklund, Nichols, & Knutsson, 2016).

For inference of whole brain searchlight data in chapter 5, a different and more sensitive approach was employed in line with searchlight procedures in recent classification studies (Hafri, Trueswell, & Epstein, 2017; Wurm, Caramazza, & Lingnau, 2017). Threshold-free cluster enhancement (TFCE; Smith & Nichols, 2009) is an alternative approach that allows for valid cluster-level inference that offers greater sensitivity in many situations than height and cluster-based thresholding (i.e. as described above). TFCE works as follows: A *non-thresholded* input image (e.g. whole brain searchlight map) is passed through an algorithm whereby the output image voxel values represent a measure of clustered 'local spatial support', or weighted sum of clustered signal in neighbouring voxels. This is calculated by iteratively increasing a signal intensity threshold (across the brain) and summing the number of connected 'supra-threshold' voxels, with the sum at lower thresholds receiving lesser weighting than at higher thresholds. As such, the resulting highest voxel values are those with the *most* local support but due to greater weighting at higher intensity thresholds *also* corresponds closely to the highest values in the raw input image (e.g. voxels with the highest searchlight classification accuracies). To correct for multiple comparisons, permutation tests are then performed to create a final z-score image (i.e. voxels represent z-scores) that is finally used for valid inference of clusters.

In the chapter 5, the following procedure was employed for each subject's searchlight maps separately. First, t-tests were performed on the searchlight map(s) and

the resulting t-values were converted to z-scores in the output 'pre-TFCE' image; for classification against chance-level classification, a one-sample t-test was performed, while a paired t-test was performed for comparing two searchlight maps. The resulting output image was then passed through the TFCE algorithm, which generated an 'actual' TFCE image (i.e. an unpermuted iteration of the TFCE procedure); this image was then compared against a further 10'000 Monte Carlo permutations of the same TFCE approach.

Permutations were achieved by randomly 'positive-flipping' or 'negative-flipping' voxel values in the 'pre-TFCE' image (e.g. a voxel value of +1.2 might become -1.2, or vice versa). Once all permutations were run, a multiple-comparison corrected map of p-values was created where each voxel value corresponds to the number of permuted z-scores (for that voxel) that exceed the corresponding z-score from the 'actual' TFCE image (e.g. 4 out of 10'000 permutations > 'actual' TFCE image = .0005; that is, the 'actual' value is the 5th highest amongst all iterations). For easy inference, this p-value map was then converted to a z-score map. Subjects' z-score maps were then averaged to create a final group level map that was thresholded at $z = 1.65$ (i.e. $p = .05$, one-tailed) for whole brain inference of above-chance classification. Therefore, the resulting z-score map shows searchlight voxels that showed reliable above-chance classification, with greater values denoting greater classification accuracy and 'locally clustered support'.

Chapter 3

Neural Responses to Visually Observed Social Interactions*

* The contents of this chapter are identical to the corresponding manuscript published in Neuropsychologia (see reference below) but has been reformatted for this thesis. The use of 'pSTS' and 'TPJ' in this chapter are contrasted with different naming conventions in chapters 4 and 5 – that is, 'pSTS-I' and 'TPJ-M', respectively; however, despite these naming differences, these regions are 'conceptually identical' across experiments.

Walbrin, J., Downing, P., & Koldewyn, K. (2018). Neural responses to visually observed social interactions. *Neuropsychologia*, *112*, 31-39.

Author contributions:
J.W. was the principle contributor for experiment 2 in this manuscript, and was centrally involved in designing the experiment, testing participants, analyzing data, and writing the manuscript.
K.K. was the sole contributor for the contents of experiment 1 (i.e. collecting and analyzing data and writing up methods and results); K.K was centrally involved in designing experiment 2 and writing and editing the manuscript.
P.D. contributed to study design and analysis of experiment 2 and with manuscript editing.

Abstract

Success in the social world requires the ability to perceive not just individuals and their actions, but pairs of people and the interactions between them. Despite the complexity of social interactions, humans are adept at interpreting those interactions they observe. Although the brain basis of this remarkable ability has remained relatively unexplored, converging functional MRI evidence suggests the posterior superior temporal sulcus (pSTS) is centrally involved. Here, we sought to determine whether this region is sensitive to both the presence of interactive information, as well as to the content of qualitatively different interactions (i.e. competition vs. cooperation). Using point-light human figure stimuli, we demonstrate that the right pSTS is maximally activated when contrasting dyadic interactions vs. dyads performing independent, non-interactive actions. We then used this task to localize the same pSTS region in an independent participant group, and tested responses to non-human moving shape stimuli (i.e. two circles' movements conveying either interactive or non-interactive behaviour). We observed significant support vector machine classification for both the presence and type of interaction (i.e. interaction vs. non-interaction, and competition vs. cooperation, respectively) in the pSTS, as well as neighbouring temporo-parietal junction (TPJ). These findings demonstrate the important role that these regions play in perceiving and understanding social interactions, and lay the foundations for further research to fully characterize interaction responses in these areas.

3.1. Introduction

Social interactions are complex and dynamic and yet can quickly convey rich information about the actions, intentions, personality and goals of the participants involved. From an early age, people use the social interactions they observe to decide who to trust, who to learn from and who is in charge. Despite the importance of our ability to parse social interactions in building our knowledge of others and the relationships between them, relatively few studies have investigated the brain's response to observed social interactions between multiple actors. Instead, the bulk of the current 'social vision' literature has focused on the perception and appraisal of individual agents. One aspect of such work has greatly increased our understanding of the social brain and has implicated a core set of regions in perceiving and evaluating individual social objects (e.g. Downing et al., 2001; Kanwisher et al., 1997). What is not yet known is whether the processing of observed social interactions is similarly supported by focal, selective regions in the human brain.

A likely candidate in which to find such a region is the posterior superior temporal sulcus (pSTS). The pSTS has been described in the literature as the 'hub' of the social brain (Lahnakoski et al., 2012) and is often included in not only the network of areas involved in 'person perception', but also in the 'action observation network' and the 'mentalizing network' (Yang et al., 2015; Deen et al., 2015). In humans, converging functional magnetic resonance imaging (fMRI) findings implicate the pSTS (as well as neighbouring posterior superior temporal gyrus), as a region that may be sensitive to visually observed social interactions. Across a number of studies, univariate pSTS responses are greater when viewing videos of interactive behaviour relative to less interactive behaviour, especially within the right hemisphere. For example, interacting dyads > individual actions (Dolcos et al., 2012; Iacoboni et al., 2004), interacting dyads > non-interacting dyads (Centelles et al., 2011), and contingent interactions > 'mirrored' interactions – that is, two agents' actions that are contingent upon each other > the same synchronised action reflected and performed by both agents (Georgescu et al., 2014). Similarly, a few recent multivariate fMRI studies have also demonstrated sensitivity to interactive behaviour within the pSTS (Hafri et al., 2017; Baldassano et al., 2017). What's more, recent evidence in macaques also implicates the STS as a central region in the visual analysis of conspecific social interactions (Sliwa & Freiwald, 2017).

Interestingly, the pSTS response to social interactions does not seem to be dependent on perceiving *human* actors in these scenarios. Indeed, similar responses have been demonstrated with simple moving shape stimuli (i.e. self-propelled geometric shapes that create robust impressions of intentional behaviour) when contrasting interactions with other forms of shape motion; for example, greater pSTS activation is observed when viewing two shapes engaged in a complex interaction relative to shapes moving in an aimless, non-intentional manner (Castelli et al., 2000; Gobbini et al., 2007; Osaka et al., 2012; Tavares et al., 2007), or similar interactions vs. non-intentional, mechanical shape movement (Martin & Weisberg, 2003).

Importantly, however, most such paradigms have not sought to directly investigate social interaction *perception* – that is, observing interactions without a task that requires explicit judgements or inferences about agents' behaviour. Across most of these studies, participants were required to make *explicit* theory-of-mind (ToM) judgements such as rating how 'intentional' shapes' movements appeared to be (Castelli et al., 2000). Whilst *implicit* intentional processes are likely evoked when simply viewing moving shape displays (e.g. understanding the immediate purpose of an action), pSTS activation is observed when individuals make explicit intentional inferences, that is, *deliberative thinking or reasoning* about the contents of an individual's mind. The following extreme case demonstrates the influence of explicit inferences in the absence of socially meaningful behaviour: Lee et al. (2012) observed increased right pSTS activity when participants made explicit intentional attributions (i.e. detecting which shape chased another) based on random shape movement compared to non-social motion judgements. At present, it is unclear whether pSTS responses to abstract depictions of social interactions are evoked by the *presence* and *contents* of the social interaction itself (i.e. whether an interaction is taking place, and what is happening in an interaction, respectively), or if previously reported responses in the pSTS are driven by differences in animate motion or the task of making explicit social judgements.

Whilst the preceding evidence demonstrates that the pSTS is sensitive to visual interactive behaviour, much remains to be learned about what this region computes about such behaviour, especially qualitatively different interactions; for example, how might interactions in which two agents compete with each other be differentiated from those in which they cooperate with each other? Previous studies have found some pSTS modulation when comparing qualitatively different interactions: Canessa et al. (2012)

used still photographic stimuli and reported only minimal right pSTS activation when contrasting cooperative > affective interactions. Sinke et al. (2010) observed increased activation in the right pSTS for threatening interactions relative to teasing interactions; however, this difference was only observed when performing an orthogonal task (i.e. attending to the colour of dots superimposed on the interaction), but not when performing an explicit inference task (e.g. identifying the emotional tone of the interaction) even though mean activation was greater during the inference task. These findings provide very preliminary evidence that the pSTS is sensitive to qualitative differences between otherwise visually similar interactions. One limitation of these studies is that they relied on univariate methods. In order to fully answer the question of whether the pSTS is sensitive to the actual *content* of an interaction rather than simply sensitive to the *presence* of an interaction, a multivariate approach may be required. Exploring content sensitivity in this fashion will provide clues to the functional role(s) of the pSTS in perceiving and understanding other people in interaction.

In experiment 1, we used point-light human stimuli across a large group of participants to demonstrate that the right pSTS is the most strongly activated region when contrasting interactions > visually matched non-interactions. In a separate group of participants, we then used the same contrast to localize human interaction-sensitive cortex within the pSTS before using a multivariate approach – that would afford greater sensitivity – to test whether this area would also contain information about the presence and qualitative content of abstract moving shape interactions. The task and stimuli were designed to focus on the role of pSTS in interaction perception while tightly controlling many 'low-level' visual cues and also attempting to reduce the influence of other social features known to engage this broad region. Specifically, these dynamic displays depicted highly-controlled moving shape stimuli that did not contain face or body information. Further, we used an orthogonal response task that did not require reporting interaction content, to minimize, as much as possible, explicit ToM inferences.

This approach allowed us to test two main hypotheses within the pSTS: Firstly, that above-chance classification of interactions vs. non-interactions would be observed. Secondly, we predicted that above-chance classification of different kinds of interactions. Specifically, we contrasted competitive vs. cooperative scenarios, as these types of interactions do not convey strongly differential emotional valence, which could confound interaction classification. A further practical advantage of testing these scenarios is that

they are easily represented with moving shape animations. In the present study, for example, we conveyed cooperation via two shapes pushing the same side of an object together, and competition via two shapes pushing opposite sides of an object.

We then tested how *anatomically specific* such a pattern of results might be by comparing pattern classification performance in the localized interaction region to a neighbouring ToM-localized region in the temporo-parietal junction (TPJ). In addition, we also included a control region in the right lateral occipito-temporal cortex (LOTC) – an area centrally involved in action perception (Lingnau & Downing, 2015) – and recently shown to robustly classify between observed actions in the context of interactions, for example, different interactive actions (Hafri et al., 2017) and interactive compared to non-interactive actions (Wurm et al., 2017). Finally, we included additional univariate analyses to compare contrasts in the pSTS and TPJ with those from a recently published study that used similar stimuli (Isik et al., 2017).

3.2. Experiment 1: Methods

Participants

Fifty participants (48 right-handed, 28 females, aged 19 – 34, *M* = 23.6 years, *SD* = 3.62) participated in the study. One participant's data was omitted from further analyses due to excessive head motion. The MIT Committee on the Use of Humans as Experimental Subjects reviewed and approved the experimental protocol and participants completed informed consent forms before taking part.

Paradigm

Participants viewed point-light dyads who either faced each other and were clearly engaged in a social interaction (e.g. both gesturing towards each other) or engaged in two independent actions (e.g. one riding a bike while the other walked). The independence of the two actions was further underscored by having the two figures face away from each other, and a line was placed down the center to form a 'wall' between the characters. The source of the interacting dyads was from Manera et al. (2010) and the source of the independent actions was from Vanrie and Verfaillie (2004). Individual

videos ranged between 3 and 8 seconds (s) in length, but were blocked together to form 16s blocks. The number of videos and the length of these videos was matched between conditions. Over the course of the scan session, 40 of the participants viewed 16 blocks of each condition, while the other 9 viewed only 8 blocks of each condition. Participants were instructed to simply maintain attention on the presented videos. A variety of other data (i.e. different data across participants) was collected in the same scan session as the currently described experiment, but will not be discussed further here.

Imaging data was acquired on a Siemens 3T MAGNETOM Tim Trio Scanner at the Athinoula A. Martinos Imaging Center at MIT using a 32-channel head coil. Functional data were collected using a T2*-weighted echo planar imaging (EPI) pulse sequence (TR = 2000ms, TE = 30ms, flip angle = 90°, FOV = 192 x 192mm, matrix = 64 x 64 mm, slice width 3mm isotropic, gap = 0.3mm, 32 near-axial slices). In addition, a high resolution T1-weighted anatomical image (multi-echo MPRAGE) was collected (TR = 2530ms, TE = 1.64ms, 3.44ms, 5.24ms, 7.014ms (combined with an RMS combination), echo spacing = 9.3ms. TI = 1400ms, flip angle = 7°, FOV = 220 x 220 mm, matrix size = 220x220mm, slice thickness = 1mm, 176 near axial slices, acceleration factor = 3, 32 reference lines).

Data Analysis

All preprocessing steps and general linear modeling (GLM) was performed using Freesurfer version 5.3 (freesurfer.net). Preprocessing consisted of standard motion correction and then the alignment of each functional run to that participant's anatomical volume. Functional data were then smoothed using a 5mm full width at half maximum (FWHM) Gaussian kernel. Smoothed data were used when defining regions of interest (ROIs), but percent signal-change data was extracted from unsmoothed data. For group level analyses, data were normalized to the Freesurfer FSAverage template and a surface-based random effects group analysis was run across all participants, weighted by the amount of data contributed by each participant (i.e. the nine participants whose contrast maps were calculated from only two runs were weighted less heavily). All ROI analyses were performed in each participant's native anatomical space.

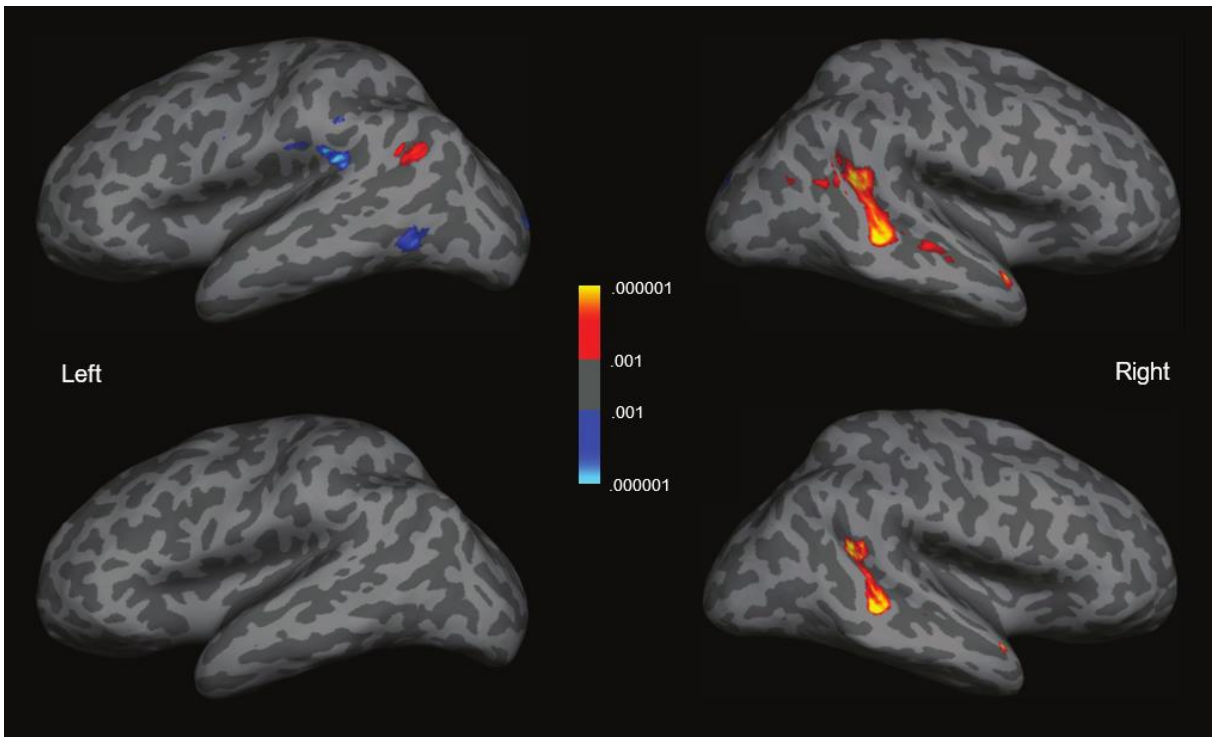


Figure 3.1. Group analysis for the social interaction > independent actions contrast (N = 49). The top panel is thresholded at $p < .001$ uncorrected while the bottom panel is FDR corrected for multiple comparisons at 5%. The colour bar represents significance (*p*-value).

ROI Definition and Analysis

Subject-specific pSTS ROIs were created using a leave-one-out method (e.g. the ROI was defined by 3 runs of data and percent signal change was calculated from the left out fourth run). This process was iterated until percent signal change had been calculated from all runs. The nine participants who had only two runs of data to work with were not included in the percent signal change analysis. ROIs were defined by intersecting an 8mm-radius sphere with the cluster peak (i.e. highest voxel t-value) in the right pSTS/STG where the contrast maps were thresholded at $p<.005$ uncorrected. Of the 40 participants entered in the ROI analysis, we could not localize an ROI for five participants, leaving 35 participants in the final ROI analysis.

3.3. Experiment 1: Results

The group analysis showed a region within the right pSTS that responded much more strongly to social interaction than to independent actions. Indeed, when false discovery rate (FDR) corrected for multiple comparisons (at 5%), only two clusters

remain – both in the STS (a large cluster in the pSTS and a small cluster in very anterior STS; see figure 3.1). Even when thresholded at a more liberal threshold of $p<.001$ *uncorrected*, this activation remains primarily in the STS, spreading more anteriorly on the right, and showing a small and much weaker response on the left (see figure 3.1). No other cortical region reached significance in this contrast. The ROI analysis revealed a significantly higher response to social interactions than independent actions (paired t-test, $t(34) = 8.20$, $p<.001$; see figure 3.2).

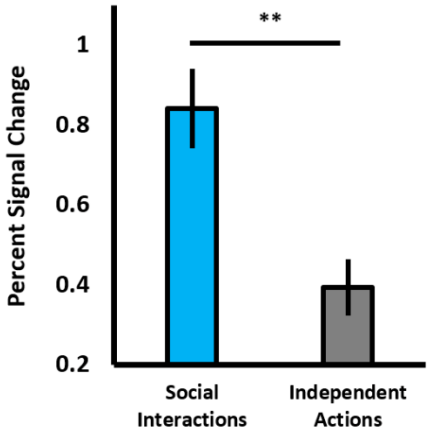


Figure 3.2. A bar chart showing mean percent signal change values for the pSTS, for both social interactions and independent actions. ** = $p \leq .001$. Error bars are SEM.

3.4. Experiment 2: Methods

Participants

23 right-handed participants (12 females; aged 18 – 30, $M = 22.43$ years, $SD = 3.07$) were recruited from the Bangor University student population. Data from two participants were removed from all analyses due to consistently low behavioral response accuracy (i.e. <50% accuracy across all runs). The study was authorized by the School of Psychology ethics committee, and participants gave informed consent and received monetary compensation for the session.

Stimuli, Design, & Task

The stimuli depicted eight different scenes from an aerial perspective (see figure 3.3), each lasting 6s. In each scene, two animate agents – blue circles – moved around a walled region in a self-propelled manner (visual angles: Agents = 0.80°; average walled space width = 7.22°). Each scenario contained a 'push-able' interaction object (e.g. a door) that served as the focus of the interaction. Eight experimental conditions were created from 2x2 factor levels: Interactive state (*interaction* & *non-interaction*) and interaction type (*competition* & *cooperation*). Interaction variants of competition and cooperation, respectively, depicted the shapes either working together, or against each other to achieve their respective or shared goals. To ensure that the outcome of each video was not confounded with the type of interaction, half of both interaction types resulted in successful object action (e.g. a door was successfully opened), and half resulted in unsuccessful action. Because successful and unsuccessful object actions had qualitatively different meanings for competition and cooperation (i.e. cooperative success results in fulfillment of both agents' action goal, whereas competitive 'success' only applies to one of the agents), we did not analyze successful vs. unsuccessful object actions. To minimize the visual familiarity of individual videos, each video was presented at four 90º rotations. In total, there were 256 novel stimuli (i.e. 4 conditions x 2 outcomes x 8 scenarios x 4 video rotations), of which 240 were randomly selected for the experiment.

As the non-interaction variants of competition and cooperation did not contain interactive behavior, the agents' movements in these conditions were generated by scrambling the agents' motion trajectories from their respective interaction conditions. This was achieved by separately splitting each agent's motion trajectory into 1s segments, and rotating the direction of each segment before joining them into continuous paths. This gave the impression of animate and self-propelled movement, but ensured that agents did not appear to interact with each other or the 'push-able' object. To reduce the appearance that agents were 'magically' causing object movement, the motion paths of the push-able interaction objects were reversed for each video to de-correlate agent and object movement. Additionally, to ensure that large differences in the proximity of agents would not drive classification differences between conditions, we ensured that agents were confined to the same region of the display as in the interaction stimuli from which they were generated (e.g. if both agents interacted within one half of the display in a given

interaction stimulus – for example, behind a closed door – we ensured that both agents' movements were restricted to the same area in the non-interaction variant of that stimulus).

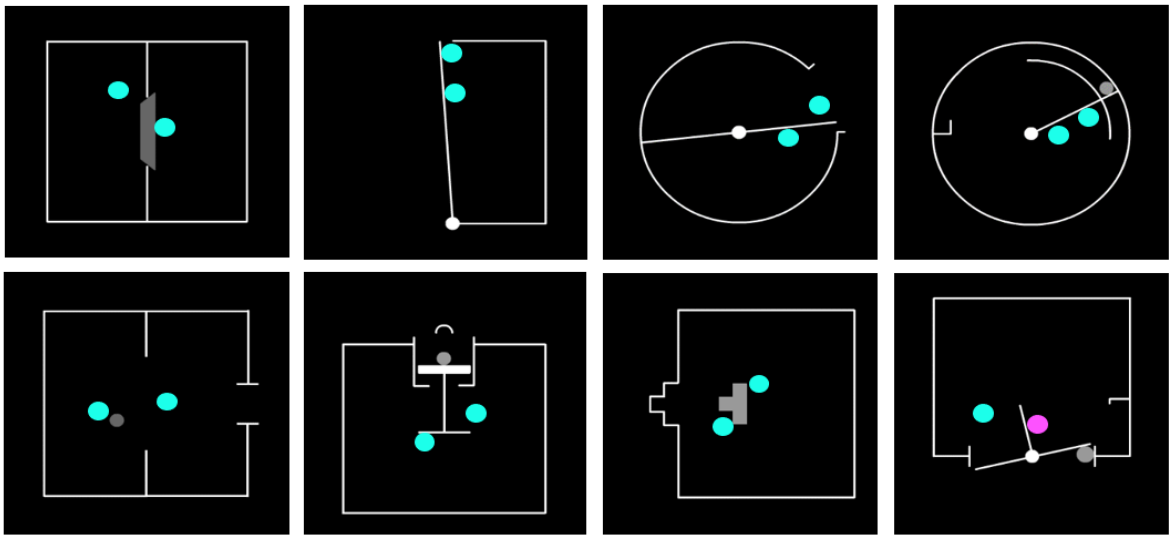


Figure 3.3. Each panel depicts one of the 8 different scenarios in which the two blue agents interacted via a 'push-able' object, as in the competition and cooperation conditions; the non-interaction conditions, by contrast, featured the same scenarios but the agents moved randomly within the scene without interacting with each other or moving the push-able object(s). The bottom right panel shows an example 'colour-change' frame in which one of the agents momentarily changed colour.

To ensure that stimulus motion energy did not differ between conditions, we used the same approach adopted in previous studies by computing differences in pixel intensity between video frames (e.g. Grezes et al., 2007). Specifically, we computed the average difference in pixel luminance between contiguous pairs of frames for each video, and then entered these values into a 2 x 2 ANOVA (i.e. interaction and non-interaction as levels of the first factor, with competition and cooperation as levels of the second factor). No term was significant (all *p*s >.462; see chapter 3 appendix D) indicating no difference in motion energy between conditions. Therefore, neural response differences between conditions could not be attributed to differences in global stimulus motion between conditions.

A blocked-design was used in which 10 runs were completed, each lasting 242s and consisted of eight experimental blocks, two per condition, 20s each (3 x 6s videos + 2s fixation epoch at the end of the block), along with three 20s rest blocks at the beginning, middle, and end of each run. A total of 80 experimental blocks were completed

(i.e. 240 experimental trials, 20 blocks per condition, 60 stimuli per condition). Block ordering was randomized across runs, and runs were randomized across participants.

Participants performed an orthogonal response task whilst viewing stimuli that was intended to minimize the tendency to make explicit ToM attributions; participants pressed a button whenever one of the agents momentarily changed color (i.e. from blue to pink). This change lasted for one animation frame (i.e. 41.67ms) and always occurred 5.5s after trial onset. To minimize the predictability of these trials, only one trial per block contained a color-change, with presentation order randomized across blocks. The brevity of the color-change also served to encourage active attending to both agents, as this change could easily be missed if a participant's attention momentarily drifted away from the agents. While not cognitively demanding, this was a relatively difficult task; it required consistent attention to both agents in order to catch the very brief change.

Localizer Stimuli, Design, & Tasks

To localize interaction-sensitive regions of the pSTS, we used a shortened version of the task used in experiment 1 (i.e. contrasting two interacting point-light figures with two figures performing independent actions). Participants completed two blocked runs (i.e. 11 x 18s blocks – four blocks per condition along with three fixation blocks, resulting in 198s run length), with blocks presented in a randomized order across runs. In order to localize ToM-sensitive cortex in the TPJ, participants also completed two runs of a written story false-belief localizer task (Dodell-Feder et al., 2011). The stimuli and protocol for this task are described elsewhere in detail (saxelab.mit.edu/superloc.php).

ROI Definition

Subject-specific pSTS ROIs were created based on activation from the independent interaction localizer task with an uncorrected height threshold of $p<.05$. ROIs were defined by intersecting a 6mm-radius sphere with the cluster peak (i.e. highest voxel t-value) in the right pSTS. As before, we chose to localize pSTS only in the right hemisphere due to observed stronger right lateralization in the group data in experiment 1, as well as in prior interaction studies (e.g. Georgescu et al., 2014). TPJ ROIs were created in an identical way (i.e. intersecting a 6mm-radius sphere with right TPJ peak activation from

the false belief > physical change contrast, with the same height threshold). Both pSTS and TPJ ROIs were successfully localized in most participants (i.e. 19 and 16 participants, respectively; see chapter 3 appendix G, Table C, for further details of these omissions). The resulting ROIs did not overlap and pSTS ROIs were observed to be significantly more anterior and ventral than TPJ ROIs across all participants in which both ROIs were localized (see chapter 3 appendix C). We defined the LOTC ROI by centering a 6mm-radius sphere at the peak coordinates (MNI x, y, z: 54 -58 -10) for the whole brain interaction > non-interaction contrast from independent data (i.e. a pilot study that used similar moving shape stimuli; see chapter 3 appendix A).

MRI Acquisition Parameters and Pre-processing

Scanning was performed with a Philips 3T scanner at Bangor University. Functional images were acquired with the following parameters: a T2*-weighted gradient-echo single-shot EPI pulse sequence; TR = 2000ms, TE = 30ms, flip angle = 90°, FOV = 230 x 230 x 132mm, acquisition matrix = 76 x 74 (reconstruction matrix = 128 x 128); 35 ascending slices (width = 3mm, gap = 0.8mm), acquired voxel size (mm) = 3.03 x 3.11 x 3.0 (reconstructed voxel size (mm) = 1.8 x 1.8 x 3.0). Four dummy scans were discarded prior to image acquisition for each run. Structural images were obtained with the following parameters: T1-weighted image acquisition using a gradient echo, multi-shot turbo field echo pulse sequence, with a five echo average; TR = 12ms, TE = 3.5ms - 10.2ms, in 1.6ms steps, acquisition time = 329 seconds, FA = 8°, FOV = 250 x 250 x 170, acquisition matrix = 252 x 224 (reconstruction matrix = 256 x 256); 170 contiguous slices, acquired voxel size = 0.99 x 1.12 x 2.0mm (reconstructed voxel size = $1mm^3$).

Before realignment, functional image runs were inspected using ArtRepair version 5b (cibsr.stanford.edu/tools/human-brain-project/artrepair-software.html**)** to assess excessive scan-to-scan head-motion. As head motion can reduce classification accuracy (Wutte et al., 2011), we adopted a strict motion threshold: Any participant runs containing >0.5mm scan-to-scan movement were omitted. Single runs were omitted from six participants, and three runs from one further participant (see chapter 3 appendix G, Table B, for full details of omissions). All pre-processing steps (i.e. realignment, co-registration, segmentation, normalization, & smoothing) were performed in SPM12 (fil.ion.ucl.ac.uk/spm/software/spm12). All default parameters were used except for a

6mm FWHM Gaussian smoothing kernel.

Data Analysis

GLM analysis was implemented on participants' normalized data in SPM12. Block duration and onsets for each experimental condition were modelled using a boxcar reference vector and convolved with a canonical hemodynamic response function. Smoothed beta maps, and subsequent t-maps, were generated for the two localizer task contrasts, whilst unsmoothed beta maps were created for classification and univariate percent signal change (PSC) analyses. Four separate regressors (i.e. competition and cooperation, along with their respective non-interaction variants) were generated for each run. For the interaction vs. non-interaction contrasts, the corresponding regressors were modelled together, that is, interaction = [competition + cooperation], non-interaction = [non-interaction competition + non-interaction cooperation].

The Decoding Toolbox (Hebart et al., 2015) was used to implement linear support vector machine (SVM) classification (hyperparameter C = 1) with a leave-one-run-out scheme; for each classification fold, voxel 'patterns' of beta-estimates for each condition, for all but one run were used as training data, and performance was tested on data from the 'left-out' run. This was performed iteratively until all runs had been tested (i.e. 10 iterations for 10 runs). For each classification contrast, a mean accuracy value was generated for each ROI, per participant. This value was based on the correct classification of each condition (e.g. competition or cooperation), averaged across classification folds. For each ROI and contrast, accuracy values were entered into one-sample t-tests against chance (i.e. 100% / 2 classes = 50% chance). An FDR correction threshold of $p \leq .01$, was determined based on 5 t-tests yielding one-sided p-values $\leq .05$ (i.e. .05 / 5 = .01). FDR corrected p-values of $\leq .025$ were determined for significant t-tests that followed PSC ANOVAs (i.e. .05 / 2 = .025).

Response accuracy (%) was calculated for each participant (i.e. a button-press response occurring <1.5s after color-change onset). Runs containing >2 inaccurate responses (i.e. <75% accuracy) were omitted from further analysis (i.e. 1 run from 2 separate participants) to ensure that only actively attended runs were included.

3.5. Experiment 2: Results

Behavioural Data

A 2 x 2 ANOVA was performed on response accuracy scores with interaction and non-interaction as levels of the first factor and competition and cooperation as levels of the second factor. Unexpectedly, a main effect between interaction and non-interaction was observed (interaction: M = 95.02, SD = 2.87; non-interaction: M = 91.70, SD = 7.03; $F(1,20) = 6.36$, p=.020), demonstrating that participants were more accurate in the interaction than non-interaction conditions. Neither the main effect between competition and cooperation (competition: M = 95.41, SD = 4.63; cooperation: M = 94.62, SD = 4.09; $F(1,20) = 0.38$, p=.543), nor the interaction term was significant ($F(1,20) = 1.40$, p=.251).

Additionally, we obtained stimulus ratings from a separate group of participants (N = 20) outside of the scanner. This served to aid our interpretation of what might drive greater response accuracy for interactions than non-interactions. Participants viewed a randomly selected subset of the original stimuli (i.e. after collapsing across the 4 rotation variants of each stimulus, there were 64 'unique' stimuli; 32 of these were presented – 8 per condition, balanced across successful and unsuccessful outcome variants). Participants gave Likert-scale ratings (i.e. 1 = strongly disagree, 7 = strongly agree) of the videos, for the following three statements: 'The agents interacted with each other' (interactivity); 'The agents were goal-directed' (goal-directedness); 'The agents were alive/animate' (animacy). Along with interactivity, we included ratings of goal-directedness and animacy as these constructs are shown to drive pSTS responses in moving shape displays (e.g. Gao et al., 2012). We then entered participant ratings into three 2 x 2 ANOVAs, one for each question (see chapter 3 appendix E, for full statistics). A main effect between interaction and non-interaction was observed for all three statements (all ps <.001), showing that interactions were perceived as more interactive, goal-directed, and animate than non-interactions. No other term was significant across any of the three ANOVAs (all ps >.100), indicating no perceived differences in interactivity, goal-directedness, or animacy between competition and cooperation (or the respective non-interaction variants of these conditions).

SVM Classification Analyses

To test our main hypotheses – that the pSTS would significantly differentiate interactions from non-interactions, and competition from cooperation – we performed SVM classification within subjects' pSTS ROIs (see figure 3.4). For both main contrasts, mean classification accuracy was significantly greater than chance (interaction vs. non-interaction: M = 65.20%, SD = 11.15, $t(18) = 5.94$, $p<.001$; competition vs. cooperation: M = 56.97%, SD = 11.15, $t(18) = 2.53$, $p=.010$). For the TPJ, we also observed above-chance classification for both contrasts, but at the uncorrected ($p\leq.05$) level only (interaction vs. non-interaction: M = 56.24%, SD = 13.07, $t(15) = 1.90$, $p=.038$; competition vs. cooperation: M = 57.35%, SD = 13.80; $t(15) = 2.13$, $p=.025$). Therefore, above-chance classification for both contrasts was observed in the pSTS and also to a weaker extent in the TPJ.

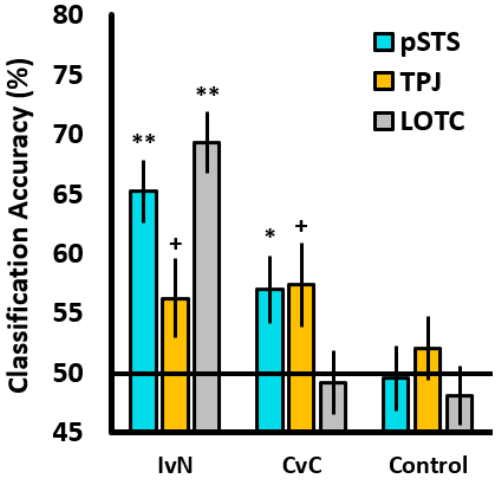


Figure 3.4. A bar chart showing mean SVM classification accuracy values (%) for each contrast. IvN = interaction vs. non-interaction; CvC = competition vs. cooperation; Control = control contrast. The horizontal line represents assumed chance-level classification (50%). ** = significant at $p \leq .001$ level; * = significant at FDR corrected level $p \leq .01$; + = significant at uncorrected $p \leq .05$ level. Error bars are SEM.

We then compared these results to a control region in the LOTC; we observed above-chance classification in the interaction vs. non-interaction contrast (M = 69.29%, SD = 11.77, $t(19) = 7.32$, $p<.001$) but not for competition vs. cooperation (M = 49.16%, SD = 11.95, n.s). These results are consistent with recent studies that demonstrate robust classification between a variety of actions, such as pushing and pulling (Hafri et al., 2017), as well as between interactive and non-interactive actions (Wurm et al., 2017); strong classification in the interaction vs. non-interaction, but not competition vs. cooperation contrast likely reflects the differential amount of object-oriented action between conditions (i.e. the agents pushed objects in *both* interaction conditions, but did not in the non-interaction conditions).

Although we found no difference in motion energy between any contrasted pair of conditions (see section 3.1.2) we sought to confirm that no low-level stimulus confounds (e.g. differences in total velocity or motion energy) might account for competition vs. cooperation classification in the pSTS and TPJ. To this end, we trained a classifier on the corresponding non-interaction contrast that was matched for motion energy (i.e. non-interaction competition vs. non-interaction cooperation; see figure 3.4) and, as expected, found no significant classification in either region (both *ps* >.466).

Univariate Analyses

In addition to classification analyses, we also sought to determine whether univariate responses differed between pSTS and TPJ, for both main contrasts. Based on the univariate contrast from experiment 1, we expected greater activation for shape interactions than non-interactions. For the competition vs. cooperation contrast however, we had no clear expectation as to whether competition or cooperation should evoke greater activation, or whether such differentiation could be captured with univariate analysis. However, a similar contrast in a recent study (Isik et al., 2017) demonstrated a trend towards greater activation for 'hindering' compared to 'helping' moving shape interactions, as well as greater activation for interactions compared with 'physical' interactions (i.e. inanimate 'billiard ball' type movements), and so we sought to determine whether our data showed similar trends.

To this end, we extracted subjects' PSC values for each condition and ROI, and ran two 2 x 2 ANOVAs – one for each of the contrasts (with ROI as the first factor, and the

respective contrast conditions as levels of the second factor). For interaction vs. non-interaction (see figure 3.5), we observed a main effect of ROI (pSTS: M = 0.43, SD = 0.42; TPJ: M = 0.05, SD = 0.30; $F(1,14)$ = 12.74, p=.003), a marginal main effect of contrast (interaction: M = 0.29, SD = 0.39; non-interaction: M = 0.20, SD = 0.34; $F(1,14)$ = 4.26, p=.058), but no interaction between the two factors ($F(1,14)$ = 1.88, p=.192). Paired t-tests (2-tailed) revealed a significant difference between interaction and non-interaction in the TPJ ($t(15)$ = 2.53, p=.023), but unexpectedly, the same trend was not significant in the pSTS ($t(18)$ = 1.59, p=.129).

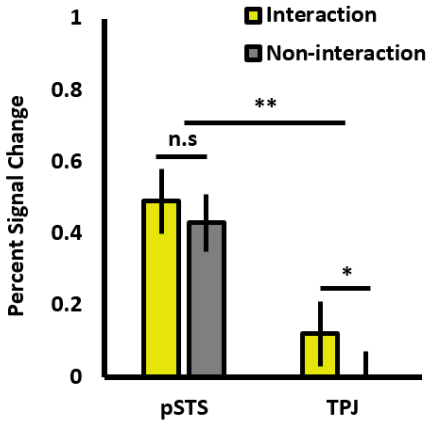


Figure 3.5. A bar chart showing mean percent signal change for the interaction and non-interaction conditions, for the pSTS and TPJ. ** = p = .003. * = significant at FDR corrected level $p \leq .025$. Error bars are SEM.

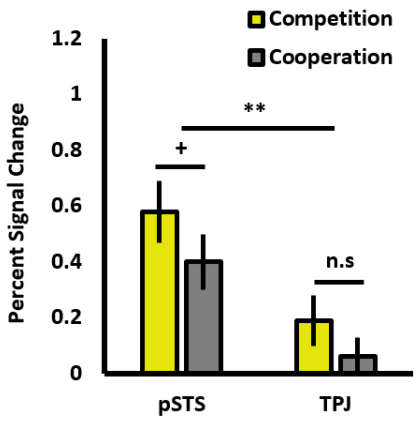


Figure 3.6. A bar chart showing mean percent signal change for the competition and cooperation conditions, for the pSTS and TPJ. ** = p = .008. * = significant at the uncorrected level $p \leq .05$. Error bars are SEM.

For the competition vs. cooperation contrast (see figure 3.6), a main effect of ROI was observed (pSTS: M = 0.49, SD = 0.41; TPJ: M = 0.12, SD = 0.32; $F(1,14)$ = 9.71, p=.008), along with a marginal main effect of contrast (competition: M = 0.37, SD = 0.48; cooperation: M = 0.21, SD = 0.35; $F(1,14)$ = 3.61, p=.078), and marginal interaction ($F(1,14)$ = 3.94, p=.067). Follow-up t-tests (2-tailed) revealed greater activation for competition than cooperation in the pSTS at the uncorrected level ($p \leq .05$) only ($t(18)$ = 2.22, p=.040), but not in the TPJ ($t(15)$ = 1.68, p=.113).

3.6. Discussion

The results from experiment 1 reveal that a region in the right pSTS is strongly responsive to dyadic social interactions; it is one of only a few regions that shows such a response in a whole brain group analysis in a relatively large sample, and it responds about twice as strongly to such interactions when compared to visually similar depictions of two individuals performing non-interactive actions. The identification of the pSTS as being uniquely responsive to human interactions is replicable in an independent set of data from our own lab (see chapter 3 appendix A) as well as in data from a recent paper using similar stimuli (Isik et al., 2017). This result is not driven by aspects of human appearance known to drive pSTS responses (e.g. Deen et al., 2015): Our stimuli contained no face information, and the two conditions each contained two point-light figures moving biologically. Similarly, in experiment 2, where our stimuli contained interactions depicted by simple shapes, we observed significant results in the region of the pSTS functionally localized with human interactions: Above-chance SVM classification was observed when contrasting interaction and non-interaction stimuli, as well as when contrasting two qualitatively different types of interaction (competition and cooperation). These findings suggest the pSTS is involved not only in recognizing social interactions, but also in assessing the meaning and content of such social interactions (i.e. differentiating between different interactions). Whilst these results might suggest a unique selectivity for social interaction perception *per se* that cannot be attributed to simple differences in face or body information, there are several caveats that prevent such a clear and unequivocal interpretation.

Firstly, the univariate results in experiment 2 do not show the same trend as in experiment 1 – that is, the interaction > non-interaction contrast for our moving shape stimuli did not reach significance in the pSTS. Secondly, we observed greater response accuracy for interaction than non-interaction moving shape stimuli in the behavioral data suggesting that, against our expectations, attentional or explicit ToM-related differences between the interaction and non-interaction conditions are likely. Thirdly, independent ratings of the moving shape stimuli showed that interaction stimuli were perceived as significantly more goal-directed and animate than non-interactions, and therefore these sources of information likely contribute to pSTS responses to these conditions. It is worth noting, however, that these potentially confounding contributions did not drive

univariate responses in the pSTS to be significantly greater for the interaction condition than the non-interactive condition in this dataset. They may be, however, a source for the difference we see in the TPJ between these two conditions. Nevertheless, given the complications in our data we cannot make any strong claims for interaction selectivity in the pSTS based on the interaction vs. non-interaction contrast in experiment 2 alone.

However, these limitations do not apply to the competition vs cooperation contrast in experiment 2. We observed significant classification and univariate results in the pSTS to this contrast that cannot be attributed to differences in response accuracy, independent stimulus ratings, or motion energy (i.e. these conditions were matched across these measures). Although there is mixed evidence as to whether the pSTS reliably differentiated the interaction > non-interaction contrast across our two stimulus sets, our univariate and classification results give direct evidence that the pSTS differentiates between qualitatively different interactions and suggest that the pSTS might play a central role in the perception and understanding of social interactions. Whilst further evidence is undoubtedly required to fully support this possibility (e.g. we cannot differentiate whether perceptual differences or potential qualitative differences in the interactors' intentions might drive this effect), the present findings align closely with a very recent paper investigating a similar research question and using conceptually similar stimuli (Isik et al., 2017). Together, these two studies suggest a central role for a region in the pSTS during social interaction perception, however the precise role(s) will need to be further specified by future research.

We also found that the pSTS was not the only region to differentiate between our conditions; similar classification responses were observed in neighboring ToM-task localized TPJ, although these responses were somewhat weaker. In addition, their relative responses to the univariate interaction > non-interaction contrast are similar, although the difference was only significant in the TPJ. It is, however, worth noting that in a similar contrast in Isik et al. (2017), as well as in pilot data we collected (see chapter 3 appendix A), the interaction > non-interaction contrast *did* reach significance in the pSTS. Conversely, for the competition vs. cooperation contrast, similar differentiation of conditions was observed in both regions, but only reached (uncorrected) significance for the pSTS. Therefore, similarities in classification performance and mixed univariate results prevent a clear functional separation of the two regions.

In some ways, the similarity between the two regions is not surprising. They occupy neighboring cortex, and functional and structural distinctions between the TPJ and pSTS are not always clear in the literature (e.g. Mars et al., 2012). In terms of overall univariate response magnitude, the two regions do respond quite differently to our shape stimuli, with the pSTS responding robustly, while the TPJ shows a relatively weak response. Indeed, much greater univariate responses in the pSTS than the TPJ were also observed in a recent study by Isik et al., (2017). Despite this univariate difference, we cannot make strong claims for functional separation of these regions based on the present data. However, the relatively stronger classification of interaction vs. non-interaction contrast in the pSTS, along with substantially greater overall univariate response to all the stimuli in the pSTS, might indicate that the two play different roles. If this is true, we speculate that the pSTS may be driven more by visual interaction cues, and the TPJ might play a relatively stronger role in processing explicit inferential information. This interpretation aligns with a trend found in a previous meta-analysis (Schurz et al., 2014): Explicit ToM inferences (e.g. false-belief inferences) that do not require visual action observation tend to activate more posterior regions of the temporo-parietal cortex, whereas tasks that involve extracting intentions from visual actions (e.g. moving shapes) evoke more anterior activation of this area (i.e. pSTS and posterior middle temporal gyrus). Similarly, prior evidence demonstrates that regions of the pSTS respond strongly to a variety of dynamic social visual information (Deen et al., 2015; Lahnakoski et al., 2012), and specifically to visual cues that underlie dynamic interactions, such as correlated motion between moving shapes (Schultz et al., 2005) and action contingencies (Georgescu et al., 2014).

However, a purely visual account the pSTS seems unlikely given top-down modulation of the pSTS when viewing moving shapes (Lee et al., 2012) and apparent sensitivity to the intentional content of visually observed actions (Pelphrey et al., 2004; Saxe et al., 2004). Instead the pSTS may play an intermediate or integrative role between perception and higher level cognition. Similarly, regions of the STS are demonstrated to integrate multimodal perceptual information (e.g. Beauchamp et al., 2004), whilst one recent study showed a fundamentally integrative neural response to holistic human-object interactions (relative to averaged responses of the constituent parts of these interactions; Baldassano et al., 2017). Understanding interactions undoubtedly requires rapid online integration of multiple dynamic actions between agents, and so the pSTS

may play a direct role in the translation of dynamic multi-agent social actions into higher level social cognition, such as understanding the immediate intentions of the interactors, and the purpose of a social encounter.

It is also important to note that the present research focuses on observed *third-person* interactions, yet these findings may inform other lines of interaction research: For example, Schilbach et al. (2013) have emphasized the importance of extending insights from research focused on third-person social scenes research to second-person interactions (i.e. measuring neural responses of an individual engaged in a real-time interaction with another individual). The extent to which the pSTS is modulated during second-person interactions remains to be fully determined, although pSTS and TPJ modulation has been observed in second-person joint attention tasks (Redcay et al., 2010; i.e. locating a visual target by following the eye gaze of an individual presented on a screen - 'joint attention' > locating the target without the others' eye gaze - 'solo attention'). By contrast, second-person interactions that emphasize *social cognition* and do not require *person-perception* (e.g. playing a computer game against an unseen opponent) typically recruit fronto-parietal regions (e.g. superior frontal gyrus and superior parietal gyrus; Decety et al., 2004) that may include mentalizing network regions such as TPJ and precuneus (Tsoi et al., 2016) but no pSTS responses are noted. Further research should aim to clarify the extent to which visual interactive cues are modulated by viewer perspective, along with the contribution of social cognition to responses in these areas.

The present data suggests that the pSTS, and similarly, the TPJ are central to interaction perception and should motivate further research to manipulate specific interaction cues to better characterize responses in these two regions. This is a particularly interesting prospect given a number of recent findings that suggest specialized processing of interactive information; for example, an observed 'inversion effect' for interacting dyads, relative to non-interacting dyads (Papeo, et al., 2017), perceptual 'chunking' of interacting dyads in working memory (Ding et al., 2017), and qualitatively richer neural representations for human-object interactions than for isolated human and object representations averaged together (Baldassano et al., 2017). It remains to be determined what the full set of cues involved in social interaction perception may be, the relative strength of and interactions between these cues, and whether the pSTS is sensitive to such cues in the absence of social information. There is

still a great deal of work to be done to build a complete model of the visual perception of interactive behavior and the brain networks that support such perception, but it is work that we believe will be both fascinating and rewarding.

Chapter 4

The Neural Development of Social Interaction Perception

Abstract

Compared to face, body, and biological motion perception, developmental investigations of social interaction perception are scarce. Recent evidence demonstrates that the posterior superior temporal sulcus (pSTS) plays an important role in recognizing and understanding third-person social interactions in adults. Although children often (but not always) show weaker cortical responses to social visual stimuli (e.g. faces, bodies, & biological motion), it is unknown whether 'adult-like' social interaction responses are present in pre-adolescent children.

This study used functional magnetic resonance imaging (fMRI) to investigate whether pre-adolescent children – relative to adults – show reduced neural responses to social interactions. Accordingly, responses to third-person social interactions were tested in 'interaction selective' pSTS (pSTS-I), along with three neighbouring regions of the posterior temporal cortex: The extrastriate body area (EBA), face selective STS (STS-F), and temporo-parietal junction (TPJ-M). Significantly greater interaction selective responses were shown for adults than children in the bilateral pSTS-I; however, this effect was marginal in the right hemisphere, demonstrating that children show weaker, but comparable responses to adults. Further analyses revealed that adults show largely bilateral responses, whereas interaction selective responses were restricted to the right pSTS-I in children. Additionally, evidence for sharper 'focal tuning' of interaction responses in both right and left pSTS-I was shown for adults, relative to weaker, more diffuse responses in children. Together, these findings show nuanced age-related differences in responses to social interactions that imply substantial developmental changes across adolescence.

4.1. Introduction

Like most cognitive processes, social visual perception undergoes substantial development across childhood and adolescence, into adulthood. Although certain aspects of face and body perception are remarkably 'adult-like' in infants and young children, such as configural processing responses (Quinn, Yahr, Kuhn, Slater, & Pascalis, 2002; Slater, Quinn, Hayes, & Brown, 2000; Zieber, Kangas, Hock, & Bhatt, 2014), other facets of these perceptual abilities – for example, sensitivity to the spacing of internal facial features, or facial emotion processing are not fully mature until later childhood or early adulthood (Batty & Taylor, 2006; Mondloch, Dobson, Parsons, & Maurer, 2004). Similarly, infants show sensitivity to human biological motion depicted by abstract 'point-light' displays, compared to scrambled presentations of point-light displays (Bertenthal, Proffitt, & Kramer, 1987; Hirai & Hiraki, 2005). However, adult-like discrimination of point-light biological motion from noise dots is not achieved until adolescence (Hadad, Maurer, & Lewis, 2011) or early adulthood (Freire, Lewis, Maurer, & Blake, 2006).

In addition to recognising and understanding social information by observing individuals, children also show some ability to extract other information, from dyadic interactions, such as determining the shared goals of two interactors. Fawcett and Gredebäck (2013) showed that, after viewing videos of two actors collaborating (contributing to a shared goal), 18 month-old infants make anticipatory eye gaze movements towards locations that denoted this shared goal in subsequent videos. However, when such sequences were preceded by videos that did not depict collaborative behaviour, infants did not make anticipatory gaze shifts, showing that infants are able to observe and infer collaborative intentions from dyadic interactions. Similarly, after viewing brief video clips of object-directed, collaborative behaviour between two individuals, 14 month-old infants stare for longer at objects associated with collaborative intent, than when non-collaborative vignettes were previously seen (Henderson & Woodward, 2011).

Along with shared goal understanding, younger children demonstrate an awareness of the relative social status of interactors. Brey and Shutts (2015) showed 3 – 6 year-olds video clips in which one interactor exhibited higher social status behaviours (e.g. open body posture and direct eye gaze) while the other showed low status

behaviours (e.g. hunched posture, varied gaze). 5 – 6 year-olds, but not 3 – 4 year-olds were able to reliably identify 'who was in charge', showing that younger children are less able to infer social status between interactors. Less developed social perception was also observed by Balas, Kanwisher, and Saxe (2012) who found that 3 – 5 year-olds were unable to correctly infer the presence of an unseen interactor from brief video clips of an individual child playing, whereas 9 year-old children show approximately adult-like performance.

Younger children also show less developed recognition of dyadic interactions when viewing visually constrained point-light stimuli. Centelles, Assaiante, Etchegoyhen, Bouvard, and Schmitz (2013) showed children (4 – 10 years) and adults videos of interacting point-light dyads and non-interacting dyads (i.e. two individuals performing separate, non-interactive or non-communicative gestures), and asked them to judge whether or not the figures in each dyad were communicating with each other. Younger children (4 – 6 years) were less accurate at performing this task overall and were also less accurate for the interaction than non-interaction condition (as were children aged 7 – 8 years). However, performance was similar to adults in 9 – 10 year-old children.

Functional magnetic resonance imaging (fMRI) responses to these point-light stimuli were also compared between children and adults in a recent study (Sapey-Triomphe et al., 2017). Children (8 – 11 years), adolescents (13 – 17 years), and adults (20+ years) viewed these stimuli and made judgements as to whether the two figures were acting together or separately. When contrasting interaction > non-interaction, all groups showed similar activations in the posterior superior temporal sulcus (pSTS), anterior and posterior regions of the middle temporal gyrus (MTG), middle occipital gyrus (MOG), inferior temporal gyrus (ITG), and inferior frontal gyrus (IFG). Although these broad whole brain findings suggest equivalent responses across groups, main effects of age group showed differences in several regions with peaks in the bilateral MTG, STG/STS, and IFG. Parametric changes in activation for this contrast revealed increasing fronto-parietal activity (e.g. middle frontal gyrus, precuneus, and inferior parietal lobule (IPL)) with increasing age; conversely, *decreasing* temporo-occipital activation (e.g. ITG, MTG, and STS) was observed with increasing age. These age-related differences suggest that adults recruit additional mentalizing resources, whereas children are more reliant on perceptual processing in temporo-occipital regions while observing social interactions.

However, several methodological limitations prevent a clear interpretation of this data. Firstly, the extent to which mentalizing or explicit inferential processing contributes to these results is unclear; it is possible that these differences arose from extraneous top-down processing driven by the social judgement response task, rather than visual responses to social interactions per se. Secondly, inter-subject variability in anatomy and the location of functional responses can lead to underestimation of whole brain group responses due to misalignment of regions and responses across subjects (Saxe, Brett, & Kanwisher, 2006); it is therefore possible that age group activation differences may be influenced by differential anatomical variability between groups.

In contrast to the findings of Sapey-Triomphe et al. (2017), other studies that used similar 'interaction > non-interaction' contrasts – with very similar point-light stimuli for which adult subjects *did not* make explicit social judgements – showed confined activation to the pSTS, a region proposed to play a central role in interaction perception (Isik, Koldewyn, Beeler, & Kanwisher, 2017; Walbrin, Downing, & Koldewyn, 2018). This region may therefore represent a central role in social interaction perception, and therefore a controlled comparison between adults in children in this area is worthy of investigation. This is especially true given the slow structural maturation of the broader STS area; longitudinal findings demonstrate that it is one of few regions to undergo continual reduction in grey matter across childhood and into early adulthood (Gogtay et al., 2004), and similar morphological changes are shown cross-sectionally with increasing age between 7 – 30 years (Mills, Lalonde, Clasen, Giedd, & Blakemore, 2014).

Previous comparisons of social visual STS responses between children and adults have revealed mixed findings. For example, STS responses to biological motion suggests highly similar responses between adults and children aged 7 – 10 years (Mosconi, Mack, McCarthy, & Pelphrey, 2005; Pelphrey, Singerman, Allison, & McCarthy, 2003a; Carter & Pelphrey, 2006; Pelphrey et al., 2003b). However, these studies measured responses in adults and children separately, and did not directly compare activations. Studies that do make direct comparisons between age groups, generally show greater functional STS responses for adults than children. For example, adults show greater selectivity for bodies (bodies > objects) than 6 – 11 year-old children (Ross, de Gelder, Crabbe, & Grosbras, 2014). Similarly, greater selectivity for faces (faces > [objects + buildings + scenes]) is shown for adults compared to adolescents (11 – 14 years) and children aged 5 – 8 years (Scherf, Behrmann, Humphreys, & Luna, 2007), and for a comparable contrast

(faces > objects) when comparing adults with 6 month-old infants (Deen et al., 2017). However, a similar comparison between adults, adolescents (12 – 16 years), and 7 – 11 year-old children did not reveal group differences for a similar contrast with static faces STS (Golarai et al., 2007) – a finding that is perhaps explained by the observation of stronger STS responses to dynamic than static faces (Pitcher, Dilks, Saxe, Triantafyllou, & Kanwisher, 2011).

Given differences in social interaction perception between adults and children, along with age-related structural and functional differences in the STS, this study aims to address the following two questions: 1) Do children show weaker interaction selective responses than adults in the pSTS? 2) How do other socially tuned temporal lobe regions – that plausibly contribute to social interaction perception – respond to social interactions? To answer these questions, a functional ROI approach was used to localize and test responses in regions sensitive to social interactions, faces, bodies, and mentalizing information in pre-adolescent children (6 – 12 years) and adults.

4.2. Methods

Participants

29 adults (mean age = 23.14 years; SD = 4.21; range = 18 - 35; 16 females) and 31 children aged between 6 – 12 years (mean age = 8.94; SD = 1.88; 13 females) took part in the experiment. Hand preference was determined using the Edinburgh Handedness Inventory (Oldfield, 1971) to ensure that all subjects were right hand dominant (due to different functional lateralization patterns observed in left-handed individuals). Adult participants gave informed consent and received monetary compensation. Children gave informed consent (along with assent from an accompanying guardian) and chose either gift vouchers or toys of equivalent value as compensation for participation. Ethical procedures were approved by the Bangor University ethics board.

MRI Tasks & Experimental Session

Three different video tasks were used to localize brain regions that are sensitive to different types of social information: A social interaction localizer, a face & body

localizer, and a mentalizing (theory-of-mind) localizer. Both the interaction and face & body localizer tasks consisted of a series of experimental blocks along with three rest blocks of the same length presented at the beginning, middle, and end of each run. The interaction localizer was almost identical to that used previously (Isik et al., 2017; Walbrin et al., 2018) and consisted of three runs of videos from three conditions: Interaction (i.e. two profile-view human point-light figures interacting with each other), non-interaction (i.e. two profile-view human point-light figures performing non-interactive actions, for example, one figure jumping, the other cycling), and scrambled interaction (i.e. average 'motion-matched' scrambled versions of the interactive stimuli where the coordinates of each point-light dot were randomly shifted to disrupt the perception of interactive or biological motion) (block length = 16s, based on three videos of variable length that summed to 16s; run length = 144s). Each run consisted of two blocks per condition – one presented in either half of each run – in randomized order with the other conditions.

The specific contrast used to localize pSTS 'interaction selective' regions of interest (ROIs) (pSTS-I) was interaction > scrambled interaction. This not only captured differences in interactive content, but also biological motion (unlike the more 'closely matched' interaction > non-interaction contrast that did not capture large differences in biological motion). This 'broader' contrast was chosen as it was more comparable to other localizer contrasts that were used here, and are typically used elsewhere (e.g. Julian, Fedorenko, Webster, & Kanwisher, 2012; faces > objects, rather than the relatively more 'socially matched' faces > bodies), and to account for the possibility that weaker interaction responses in children may have resulted in poorer localization of pSTS-I ROIs (i.e. children's responses may have been weaker).

The face & body localizer was adapted from stimuli used previously (Pitcher et al., 2011), and served to localize face selective cortex in the STS (STS-F) and fusiform face area (FFA), along with body selective extrastriate body area (EBA). Participants completed three runs that contained blocks depicting either moving faces, moving bodies, and moving objects (STS-F & FFA localization contrast = faces > objects; EBA localization contrast = bodies > objects; block length = 18s (6 x 3s videos); run length = 270s). Each run consisted of four blocks per condition – two presented in either half of each run – in randomized order with the other conditions.

Finally, to localize mentalizing selective temporo-parietal cortex (TPJ-M), participants viewed the Pixar short-film 'Partly Cloudy' (2009; duration = 355s, including 10s rest). Previous research (Richardson, Lisandrelli, Riobueno-Naylor, & Saxe, 2018) used a reverse-correlation analysis to identify time points within the video that reliably evoked responses to mentalizing and pain (along with 'social' and 'control' time-points, not analysed here), allowing for the localization of mentalizing and pain network regions (total time-points per condition: Mentalizing = 44s; pain = 26s ; social = 28s; control = 24s). Specifically, bilateral TPJ-M was localized by contrasting responses during mentalizing timepoints with pain timepoints (i.e. mentalizing > pain).

All scans were acquired in one session that lasted approximately 90 minutes. The scanning session was split into two halves with a short break where subjects came out of the scanner for approximately 5 – 10 minutes. This was intended to minimize fatigue in children, but for consistency, adults also took this break. Each participant completed scanner tasks in the same order, as follows: One mentalizing localizer run, three interaction localizer runs, first structural T1 scan, break outside of the scanner, second structural T1 scan, DTI scan (not analysed here), and three face & body localizer runs. Additionally, adults completed three runs of a biological motion localizer (not analysed here) that was completed before the first structural scan, in the first session half.

Prior to entering the scanner, children (but not adults) completed a short head motion 'training session'. This entailed lying still inside a 'mock scanner' with a motion sensitive electrode placed on the forehead to measure movement across three translation and three rotation axes (MoTrak Head Motion Tracking System; Psychology Software Tools, 2017). Children viewed a monitor through a mirror attached to an MRI head coil, allowing for visual motion feedback (i.e. an on-screen cursor that was 'controlled' by head movements was visible to subjects, as well as researchers via a separate monitor). Pre-recorded audio of an fMRI acquisition sequence was also played during this session, to further simulate the real scanner environment.

Subjects were instructed to lie still and 'keep the cursor in the middle of a target circle' that was presented centrally on screen. The diameter of this circle corresponded to a region allowing for 3mm head translations in any direction. Once participants were able to keep the cursor within this target region for a timed period of 30 seconds, a different task was performed without visual motion feedback; subjects were simply instructed to stay relaxed and keep still while watching a short animated video but were

told that if they moved 'too much' (i.e. > 3mm translation movements from a given initial head position), video playback would be paused, signalling that they had moved too much. Once they were able to watch the video for a period of 2 minutes without a video pause, they were deemed 'ready' to be scanned. Head movement data was not recorded or analysed, as the sole purpose of this session was to train children to remain still inside the scanner.

MRI Parameters

Scanning was performed with a Philips 3T scanner at Bangor University. The same fMRI parameters were used for all localizer tasks as follows: T2*-weighted gradient-echo single-shot EPI pulse sequence (with SofTone noise reduction); TR = 2000ms, TE = 30ms, flip angle = 83°, FOV(mm) = 240 x 240 x 112, acquisition matrix = 80 x 78 (reconstruction matrix = 80); 32 contiguous axial slices in ascending order, acquired voxel size (mm) = 3 x 3 x 3.5 (reconstructed voxel size = $3mm^3$). Four dummy scans were discarded prior to image acquisition for each run. Structural images were obtained with the following parameters: T1-weighted image acquisition using a gradient echo, multi-shot turbo field echo pulse sequence, with a five echo average; TR = 12ms, average TE = 3.4ms, in 1.7ms steps, total acquisition time = 136 seconds, FA = 8°, FOV = 240 x 240, acquisition matrix = 240 x 224 (reconstruction matrix = 240); 128 contiguous axial slices, acquired voxel size(mm) = 1.0 x 1.07 x 2.0 (reconstructed voxel size = $1mm^3$).

fMRI Pre-processing & GLM Estimation

Pre-processing was performed in SPM12 (fil.ion.ucl.ac.uk/spm/software/spm12) on all fMRI data from each session-half, separately (i.e. mentalizing and interaction localizer data with the first T1 image, and dynamic localizer data with the second T1 image). The following pre-processing steps were run: Realignment (and reslicing), co-registration, segmentation, normalization, and smoothing). All default SPM12 parameters were used except for a 6mm FWHM Gaussian smoothing kernel. General linear model (GLM) analysis was also performed in SPM12 on participants' normalized images (both smoothed and unsmoothed versions) for each localizer task separately. Event durations and onsets for each experimental condition (per run) were modelled

using a boxcar reference vector and convolved with a canonical hemodynamic response function (without time or dispersion derivatives), with a high-pass filter of 128s and autoregressive AR(1) model. Rest periods were implicitly modelled and the six motion parameters were modelled as nuisance regressors, and used in subsequent head motion analyses.

ROI Definition and Percent Signal Change Extraction

A group-constrained ROI definition procedure (Julian, Fedorenko, Webster, & Kanwisher, 2012) was implemented, and ensured that ROI definition and percent signal change (PSC) extraction were based on independent data, and is described as follows. Firstly, for each given localization contrast (e.g. interaction > scrambled interaction), for each subject, a subject-specific ROI 'search sphere' was created by running a whole brain analysis for N-1 group subjects; that is, for a given subject, *all other group subjects'* data (e.g. all other adult group data was used to create the search sphere for a given adult) was used to find the most activate voxel (i.e. highest t-value) at which an 8mm-radius sphere was placed. This sphere size was chosen to ensure minimal overlap between search spheres for different regions (e.g. pSTS-I and STS-F) so that ROIs were comprised of distinct voxels (e.g. voxels did not overlap between pSTS-I and STS-F). Minor overlap was removed between search spaces, and mean search space size was calculated for each region to ensure that the resulting search spheres were relatively similar in size (see chapter 4 appendix A).

Subject-specific search spaces were used to create the final set of ROIs. Specifically, the 100 most active contiguous voxels within these regions were selected (i.e. highest t-values for the same contrast as used to define the corresponding search space). The rationale for using 100 voxels was based on a compromise between using a much smaller number of voxels (e.g. 20 voxels) that might have resulted in exaggeratedly high selectivity values, and the maximum size of voxels within the search space (e.g. 200+ voxels). However, 9 other similar sets of ROIs based on the highest 20, 40, 60, 80, 120, 140, 160, 180 and 200 voxels were also generated for an exploratory analysis that is described below.

For the interaction and face & body localizers (for which there were three runs of data), a leave-one-run-out (LORO) approach was used to ensure that data used to define

ROIs was independent of that used to extract PSC responses. For example, an ROI created from the first two runs of data for a given contrast was used to extract data from the third 'left-out' run of data. This was iterated with each run of data designated as the 'extraction' run and the remaining runs of data used to define that version of an ROI. Subject-wise PSC extraction for each condition within each left-out run was performed in MarsBaR (Brett, Anton, Valabregue, & Poline, 2002) and the resulting values were averaged across all LORO versions of each ROI. A similar approach was used for mentalizing localizer responses, except that given only one run of data, extraction and definition were not independent of each other; however, mentalizing responses were not intended to be measured in this region and extraction of interaction and face & body localizer conditions remained independent. It is important to note that for each run of data for the interaction and face & body tasks, responses were extracted and averaged, in an identical manner as for other ROIs.

PSC and Selectivity Analyses

Mean PSC values were extracted for each subject, within each group, for all tasks (10 conditions total = 3 + 3 + 4 for interaction, face & body, and mentalizing localizer tasks, respectively) and all 9 ROIs (4 bilateral ROIs and right FFA; left FFA could not be localized due to very weak responses across subjects). The main measure of interest – that is, *selectivity* – was calculated by subtracting PSC values for the four *selectivity contrasts* as follows: *Interaction* – non-interaction; *faces* – objects; *bodies* – objects; *mentalizing* – pain. However, prior to calculating these values, a series of one-sample t-tests were performed to determine which conditions were significantly 'above-zero', within each ROI. Above-zero responses (i.e. PSC values that were significantly greater than zero) for a given target condition (e.g. interaction) were considered a pre-requisite for calculating selectivity scores as any region that was not univariately driven by a given selective category, could not be subsequently interpreted as meaningfully selective. Therefore, selectivity scores were not calculated for contrasts in regions that did not show above-zero responses for given target conditions (see table 4.1 for details of these omissions). Importantly, TPJ-M ROIs in both hemispheres were excluded from selectivity analyses as the mentalizing condition was the only target condition to show above zero

responses in both groups in these regions (although children showed some sensitivity to faces in these regions).

To assess whether interaction selectivity differed as a function of ROI size, interaction selectivity measures were extracted from 9 additional sets of pSTS-I ROIs. A linear regression slope analysis was performed on these measures as follows: For each subject, within each ROI, separately, a regression slope (beta coefficient) was calculated based on the proportional change in selectivity with increasing ROI size. As such, a linear reduction in selectivity with greater ROI size would result in a negative beta coefficient, while approximately 'fixed' selectivity across ROI size would yield a coefficient close to zero. Subjects' beta coefficients were then entered into ANOVAs and t-tests for statistical inference.

Finally, Bonferroni correction for multiple comparisons was implemented for each 'set' of analyses separately (but not for exploratory analyses); the corrected Bonferroni threshold (α) is stated before for each series of tests in the results section. All tests survived Bonferroni correction, unless otherwise stated. All one-sample t-test *p*-values are one-tailed.

Whole Brain Analyses

For group level whole brain analyses, contrast images (e.g. interaction – non-interaction) from each age group (separately) were entered into one-sample t-tests in SPM12. Similarly, for the adult > children contrast (and the reverse comparison) contrast images were entered into an independent t-test. The resulting t-maps were height-thresholded at $p = .001$, and false discovery rate (FDR) cluster corrected at $p < .05$.

Head Motion Analysis

In line with previous studies (e.g. Peelen, Glaser, Vuilleumier, & Eliez, 2009), we excluded runs of data with > 2mm scan-to-scan movement across runs. This resulted in the exclusion of single runs of data in three separate children (however including these runs did not meaningfully change group results). LORO ROI definition and extraction was therefore based on just two runs of data in these three subjects, for the corresponding tasks only (all others contained all available runs). In addition to removing runs with >

2mm movements, differences in head motion between groups were tested using a similar analysis to Kang, Burgund, Lugar, Peterson, and Schlaggar (2003). That is, a root mean square (RMS) measure of head motion across the six motion parameters was calculated per run, for each subject. This measure was also calculated for all time points of each condition separately. Both of these measures were then averaged across runs, allowing for the calculation of either 'task-wise' motion measures – for example, average head motion across all three interaction localizer runs, across all conditions – or 'condition-wise' measures (e.g. average head motion for the interaction condition across all three localizer runs).

4.3. Results

Head Motion Analysis

An initial head motion analysis revealed that, consistent with previous studies (e.g. Kang et al., 2003), children moved more than adults (see chapter 4 appendix B). Although differential head motion could confound any resulting group differences, these effects were likely to be small as we ensured that only runs of data with < 2mm of scan-to-scan motion were retained for analyses. Also, realignment in SPM is robust to considerably high thresholds of motion (Ardekani, Bachman, & Helpern, 2001). Nevertheless, a further head motion analysis was performed to determine whether potential group differences in the main measure of interest – interaction selectivity – could be attributable to group differences in head motion; specifically, interaction selectivity was calculated by subtracting non-interaction PSC responses from the interaction PSC responses. Therefore, if no head motion differences emerged between these conditions, between groups, neural interaction selectivity differences between age groups could not be attributed to differences in head motion.

To test this possibility for interaction selectivity (and other categories of selectivity), RMS head motion values across all time points belonging to a given condition (e.g. interaction), were compared with those from the corresponding condition used to calculate selectivity. For interaction and non-interaction RMS motion values, a 2 x 2 mixed ANOVA (condition x age group) was performed. Crucially, this did not reveal a main effect of condition ($F(1,58) = 0.00$, $p = .968$, $\eta_p^2 = .000$). A significant main effect of

group ($F(1,58) = 13.50$, $p = .001$, $\eta_p^2 = .189$) but no interaction between factors was also shown ($F(1,58) = 0.31$, $p = .580$, $\eta_p^2 = .005$). Therefore, head motion differences could not account for any group comparisons based on interaction selectivity, despite greater overall movement for children than adults. Similar analyses were run for the faces > objects, and bodies > objects selectivity contrast pairs; as for interaction selectivity, no confounding effects of head motion were observed for these contrast pairs (see chapter 4 appendix C for full statistics along with mentalizing vs. pain comparison).

Initial PSC Analyses

Before addressing the primary hypothesis, a 2x2 mixed ANOVA (condition x age group) was run with PSC values for each pSTS-I ROI, separately (Bonferroni corrected $\alpha = .025$). Specifically, this aimed to test for group differences between the interaction and non-interaction conditions.

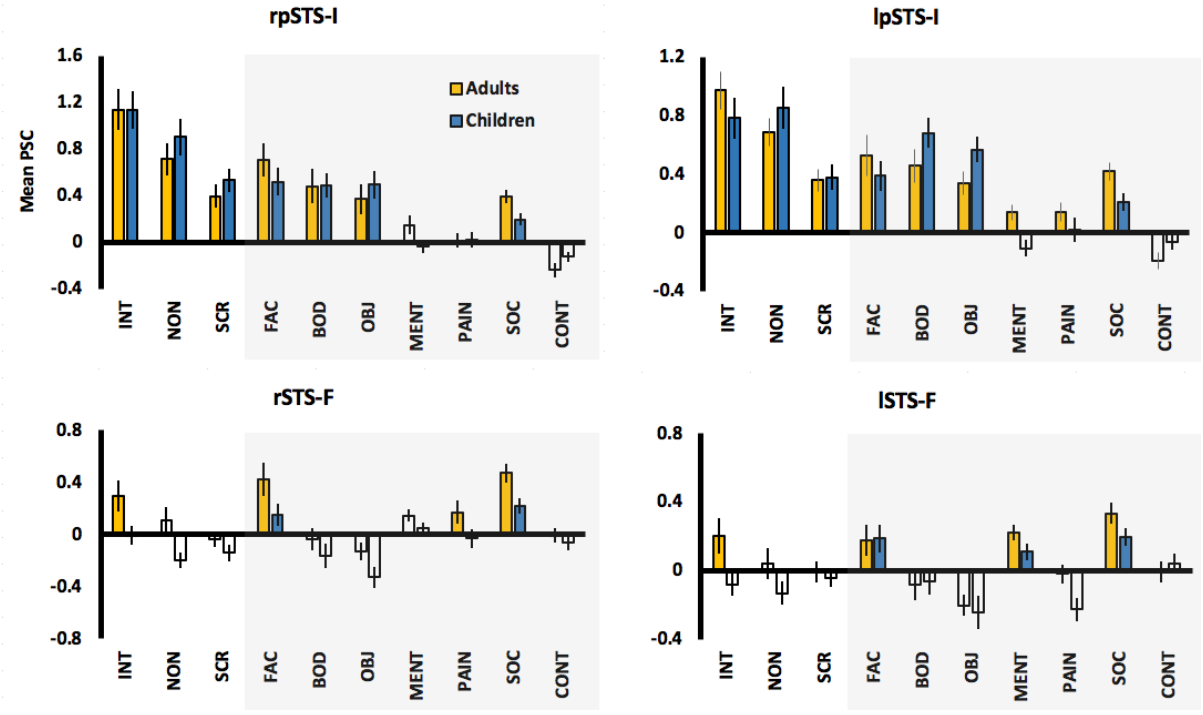


Figure 4.1. Mean PSC for the 3 interaction localizer conditions (unshaded region) along with other conditions (shaded region) in the right and left pSTS-I ROIs. Bilateral STS-F ROIs are also shown for comparison. White bars correspond to PSC values that were not significantly greater than zero (i.e. one-sample *p* value > .05). Error bars are SEM. INT = interaction; NON = non-interaction; SCR = scrambled interaction; FAC = faces; BOD = bodies; OBJ = objects; MENT = mentalizing; PAIN = pain; SOC = 'social'; CONT = 'control'.

For the right pSTS-I (see figure 4.1), greater responses were observed in the interaction compared to non-interaction condition ($F(1,58) = 34.27$, $p < .001$, $\eta_p^2 = .371$). No main effect of group ($F(1,58) = 0.18$, $p = .671$, $\eta_p^2 = .003$) was observed, and the interaction between factors was marginally significant ($F(1,58) = 2.83$, $p = .098$, $\eta_p^2 = .046$) showing a small to medium effect size. This demonstrated that, children's responses to interaction conditions in the right pSTS-I are relatively adult-like, although a trend towards group differences exists.

In the left pSTS-I, a similar pattern of results was observed – a main effect of condition ($F(1,58) = 4.47$, $p = .039$, $\eta_p^2 = .072$; although this did not survive Bonferroni correction), no main effect of group ($F(1,58) = 0.01$, $p = .944$, $\eta_p^2 = .000$), but a significant interaction was shown ($F(1,58) = 12.27$, $p = .001$, $\eta_p^2 = .175$). This strong interaction suggests nuanced group differences for the interaction localizer conditions in the left pSTS-I.

	Adults				**Children**			
	I	F	B	M	I	F	B	M
rpSTS-I								
rSTS-F								
rEBA								
rTPJ-M								
rFFA								
lpSTS-I								
lSTS-F								
lEBA								
lTPJ-M								

Table 4.1. (Above) A table showing which selectivity measures were not calculated (red cells), for each ROI, due to non-above-zero PSC values for the main localizer conditions. I = interaction; F = faces; B = bodies; M = mentalizing.

		M	**SD**	**t**	**df**	**p**
rpSTS-I	Adults	0.42	0.46	4.97	28	<.001
	Children	0.23	0.41	3.18	30	.003
lpSTS-I	Adults	0.29	0.45	3.46	28	.002
	Children	-0.07	0.34	-1.16	30	.257
rSTS-F	Adults	0.19	0.30	3.29	27	.003
	Children	0.19	0.28	3.56	28	.001
lSTS-F	Adults	0.16	0.28	3.06	27	.005
	Children	0.05	0.29	0.95	28	.351
rEBA	Adults	0.02	0.35	0.34	27	.738
	Children	0.00	0.33	0.01	28	.993
lEBA	Adults	0.14	0.38	2.02	27	.054
	Children	0.04	0.39	0.57	28	.575
rFFA	Adults	0.03	0.26	0.62	27	.540
	Children	0.00	0.36	0.07	28	.945

Table 4.2. (Right) A table showing one-sample descriptive statistics for interaction selectivity measures in the bilateral pSTS-I, STS-F, EBA, and right FFA.

Informal inspection of PSC values in both pSTS-I ROIs showed stronger responses to the interaction condition than face, body, or mentalizing conditions, and also that responses were stronger for the interaction condition in the right than left pSTS-I, for both groups (see chapter 4 appendix D). Additionally, no other region showed strong evidence for differentiation between interaction and non-interaction conditions except for bilateral STS-F ROIs (see figure 4.1; see chapter 4 appendix E for PSC charts for other ROIs). Additionally, these findings show that PSC responses in children are not uniformly weaker than adults; indeed, children showed approximately equivalent or greater PSC values than adults for many conditions across ROIs. Therefore, any differences in selectivity are not attributable to 'generally' weaker responses in children.

Interaction Selectivity in the pSTS

A series of analyses were performed to test interaction selectivity in the pSTS-I, along with other posterior temporal lobe social brain regions. Regions for which selectivity measures were not calculated (due to non-above-zero PSC values for a given 'target' condition, that is, either interaction, faces, bodies, or mentalizing) are shown in table 4.1.

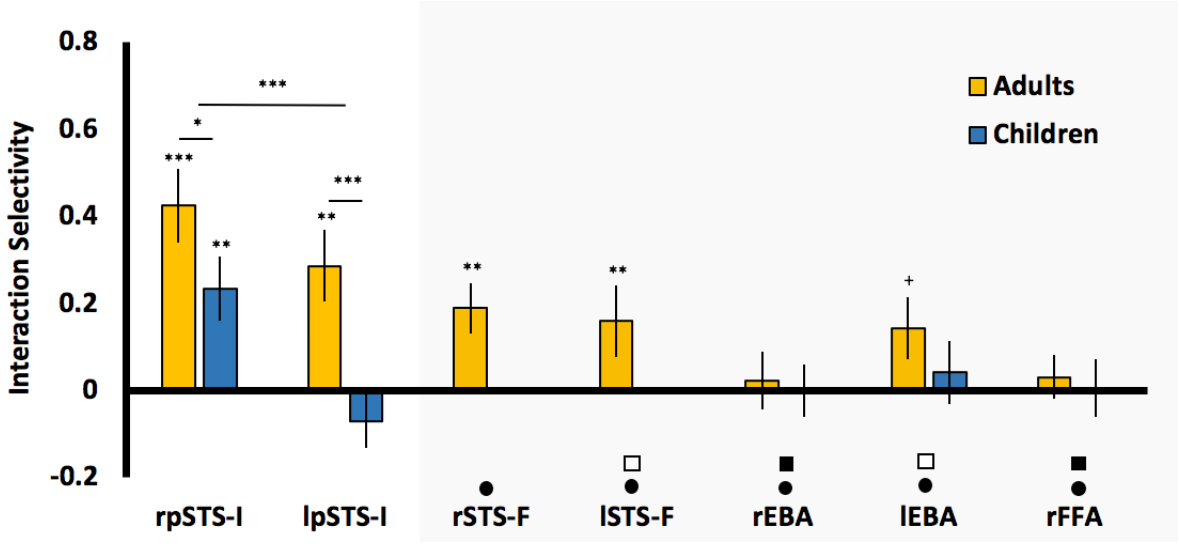


Figure 4.2. Mean interaction selectivity values for the pSTS-I ROIs (unshaded region) and other ROIs (shaded region). *** = $p < .001$; ** = $p < .01$; * = $p = .049$; $^{+}$ = $p = .054$. Error bars are SEM. Black circles denote regions that are significantly less selective than right pSTS-I for both age groups. Black and white squares, respectively, denote regions that are either significantly or marginally less selective than the left pSTS-I for adults.

One-sample t-tests (uncorrected) were performed to determine which regions showed above-chance selectivity for interactions (see table 4.2 for one-sample statistics for interaction selectivity values). For adults, strong interaction selectivity was shown for the right pSTS-I ($p < .001$), along with selective responses in the left pSTS, and bilateral STS-F (all $ps < .003$), and marginal selectivity in the left EBA ($p = .054$). By contrast, the right pSTS-I was the only significantly selective region in children ($p = .002$).

To test the main hypothesis – that adults are more interaction selective than children in the pSTS-I – two independent t-tests were performed on selectivity values in right and left pSTS-I, separately (see figure 4.2; Bonferroni corrected $\alpha = .025$). In line with the interaction terms in the initial PSC ANOVAs, adults were significantly more selective than children in the left hemisphere ($t(58) = 3.50$, $p = .001$). This effect was also observed in the right hemisphere, however this did not survive multiple comparison correction ($t(58) = 1.68$, $p = .049$). These findings show greater interaction selectivity for adults than children in the left pSTS-I, with a similar, marginal trend in the right pSTS-I. Additionally, although both groups show selective responses in the right pSTS-I, unlike adults, children do not show this trend in left pSTS-I or neighbouring bilateral face regions of the STS.

Right pSTS-I is the Most Interaction Selective ROI

Next, the prediction that interaction selectivity is greater in the pSTS-I than all other ROIs was tested (Bonferroni corrected $\alpha = .006$). First right pSTS-I and left pSTS-I were compared with a 2 x 2 mixed ANOVA (region x age group). Greater responses were observed in the right pSTS-I ($F(1,58) = 21.70$, $p < .001$, $\eta_p^2 = .272$), and greater selectivity overall for adults was shown ($F(1,58) = 8.05$, $p = .006$, $\eta_p^2 = .122$), and a marginal interaction, suggesting a trend for greater selectivity in adults compared to children ($F(1,58) = 3.11$, $p = .083$, $\eta_p^2 = .051$). Follow-up paired t-tests showed that the trend for greater right pSTS-I selectivity was statistically stronger in children ($t(30) = 4.83$, $p < .001$) than adults ($t(28) = 1.93$, $p = .032$; significant at uncorrected level only). These findings demonstrate strongly right lateralized interaction selectivity in children; by contrast, adults show more bilateral pSTS-I selectivity, with a trend towards right lateralized responses.

Further comparisons between right pSTS-I selectivity and all other ROIs were then performed (see chapter 4 appendix F for full statistics). Interaction selectivity in the right pSTS-I was significantly greater than ROIs for which *both groups* showed above-zero selectivity: The right EBA, left EBA, and right FFA (main effect of region: all *ps* < .001) with no group differences (main effect of group: all ps > .090; interaction terms: all *ps* > .080). Because adults – but not children – showed above-zero selectivity in the bilateral STS-F ROIs, two paired t-tests were performed for adults only; right pSTS-I selectivity was significantly greater than the left STS-F (p = .002), and at an uncorrected level in the right STS-F (p = .008). Therefore, in both age groups, interaction selectivity was significantly greater in the right pSTS-I than all other ROIs, although this trend was weaker against the right STS-F in adults.

Left pSTS-I Interaction Selectivity vs. Other ROIs

To test whether similar trends were true for left pSTS-I in adults (but not children as they did not show above-zero selectivity in this ROI) paired t-tests were performed between this ROI and each other region (see chapter 4 appendix F for full statistics; (Bonferroni corrected α = .01). These tests revealed that interaction selectivity was greater in the left pSTS-I than the right FFA (p = .001) and right EBA (p = .003), and greater than the left EBA (p = .027) and left STS-F (p = .044) at an uncorrected threshold. However, the trend towards greater selectivity than the right STS-F was not significant (p = .113). These results demonstrate greater interaction selectivity for adults in left pSTS-I than all other regions except right STS-F.

Interaction vs. Face & Body Selectivity in the pSTS

It was shown that interaction selectivity is greater in the pSTS-I than virtually all other ROIs, but is this region more selective for interactions than faces or bodies? To test this possibility, a 3 x 2 mixed ANOVA (selectivity category x age group) with follow-up tests in the right pSTS-I was performed, along with t-tests in adults for the left pSTS-I (Bonferroni corrected α = .01). In the right pSTS-I, a main effect of selectivity category (F(1.46,80.45) = 8.11, p = .001, η_p^2 = .129), and a main effect of age group at an uncorrected level was shown (F(1,55) = 4.92, p = .031, η_p^2 = .082). As the interaction term

was not significant ($F(1.46,80.45) = 1.05$, $p = .337$, $\eta_p^2 = .019$), follow-up paired t-tests were performed on both adults and children's selectivity scores together. Interaction selectivity was significantly greater than body selectivity ($t(56) = 4.17$, $p < .001$), however the trend for greater interaction selectivity than face selectivity was only significant at an uncorrected threshold ($t(56) = 1.75$, $p = .043$).

For adults in the left pSTS-I, greater interaction selectivity than body selectivity was observed at an uncorrected threshold ($t(27) = 2.12$, $p < .022$), but not for face selectivity ($t(27) = 1.12$, $p = .132$). These results demonstrate that, for both adults and children alike, interaction selectivity in the right pSTS-I is greater than body selectivity, and marginally greater than face selectivity in the right pSTS-I. These effects are considerably weaker for adults in the left pSTS-I (and absent for children).

pSTS Interaction Selectivity as a Function of ROI Size

The preceding analyses show greater interaction selectivity in the pSTS-I for adults than children; this difference was pronounced in the left pSTS-I, but marginal in the right pSTS-I. These selectivity measures were generated from ROIs with a fixed size of 100 voxels, and therefore ensured that no group differences in ROI size existed. However, without knowing the underlying functional organization of selective responses within the pSTS – and crucially, whether this differed between groups – it is possible that the current choice of ROI size might have unwittingly favoured adult responses. For example, selecting the highest 100 contiguous voxels in adults might approximately capture the entire peak of an interaction selective cluster, whereas this peak region might be smaller in children, and as such, selectivity measures in children may have been

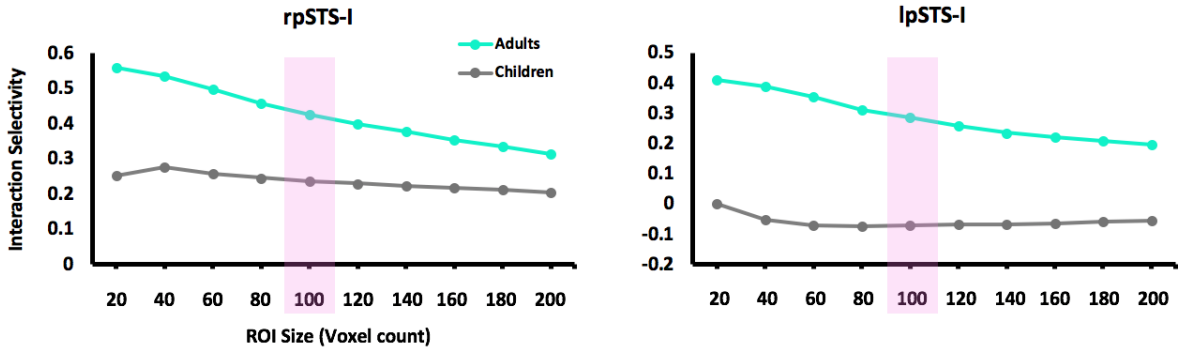


Figure 4.3. Mean interaction selectivity plotted as a function of ROI size, for both age groups, in the right and left pSTS-I. The shaded region shows the 100 voxel ROI, as used in the preceding analyses.

calculated with the inclusion of less selective voxels outside of the peak). An exploratory analysis was conducted to determine if selectivity differed as a function of ROI size, and whether such changes were similar between groups.

Interaction selectivity as a function of ROI size is shown in figure 4.3**.** This shows greater selectivity for adults than children across all ROI sizes. Interestingly, adults show a trend towards linear decreases in selectivity with increasing ROI size, whereas this does not appear to be the case for children. To test whether the apparent age group differences in selectivity changes were significant – that is, that selectivity linearly decreases as a function of ROI size in adults, but not children – beta coefficients (linear regression slopes) were extracted for each subject, per ROI and entered into a series of group tests (uncorrected). One-sample t-tests revealed that adult beta coefficients were significantly lower than zero in both the right (M = -0.0015; SD = 0.0032; $t(28)$ = -2.58, p = .007) and left hemisphere (M = -0.0015; SD = 0.0026; $t(28)$ = -3.09, p = .002), but this was not true for children (right pSTS-I: M = -0.0003; SD = 0.0015; $t(30)$ = -1.30, p = .103; left pSTS-I: M = -0.0003; SD = 0.0015; $t(30)$ = -0.91, p = .373). A 2 x 2 mixed ANOVA (hemisphere x age group) was performed to directly confirm if the two groups differed, and whether this difference was consistent across hemispheres. A main effect of age group was observed ($F(1,58)$ = 7.05, p = .010, η_p^2 = .108), but neither the main effect of hemisphere ($F(1,58)$ = 0.01, p = .912, η_p^2 = .000), nor the interaction term was significant ($F(1,58)$ = 0.00, p = .992, η_p^2 = .000). These findings show that interaction selectivity decreases linearly with increasing pSTS-I ROI size in adults, but not children, in both hemispheres. It is speculated that this might indirectly reflect age-related differences in the *focal tuning* of interaction responses in the pSTS; that is, adults show strongest selectivity around a small peak in the pSTS, but children are less selective and more broadly tuned to interactions in the right pSTS only.

Whole Brain Analysis: Interaction > Non-interaction

To determine whether other regions outside of the functionally localized ROIs demonstrated sensitivity for the interaction > non-interaction contrast, whole brain analyses were performed (see figure 4.4). For adults, strongly right lateralized responses were shown with peak activation in the pSTS, along with activations extending to the anterior STS (aSTS), and a small cluster in IFG. Additionally, bilateral precuneus and small

left calcarine sulcus activation was shown. For children, similar although weaker responses were shown in the right pSTS and aSTS regions only.

When comparing activation between the two groups directly, no differences were observed for either the adults > children, or the reverse contrast. In summary, no statistically significant differences emerged between groups for the interaction > non-interaction contrast. Instead both groups showed right lateralized responses in the pSTS as well as aSTS.

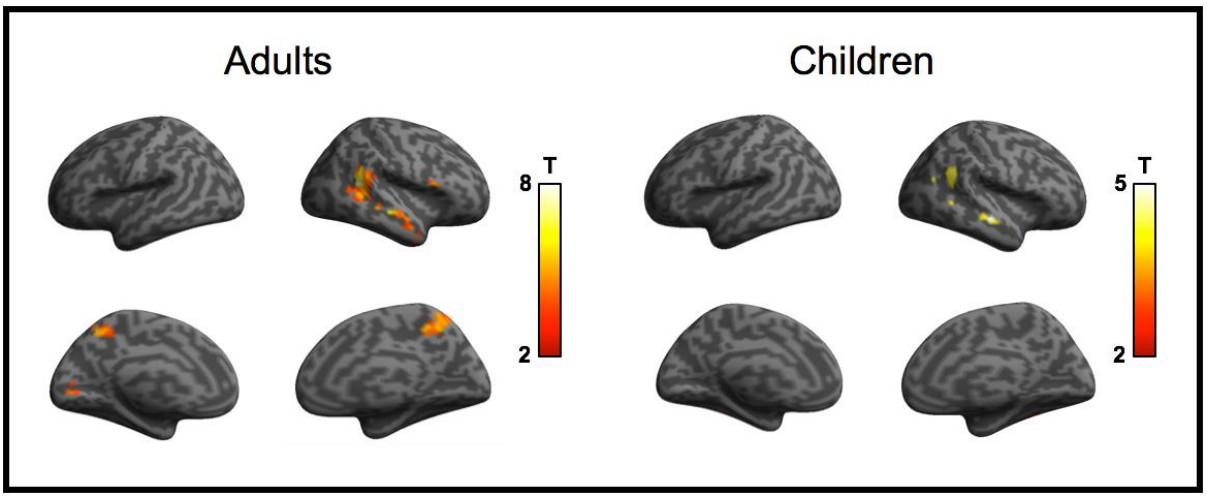


Figure 4.4. Whole brain activation for the interaction > non-interaction contrast for each age group. Colour bar represents activation t-value, scaled for each group, separately. Height threshold = .001; FDR cluster correction = .05.

4.4. Discussion

Results Overview

This study aimed to investigate potential differences in neural responses to social interactions between adults and children – both within interaction selective pSTS and other socially selective regions of the posterior temporal lobe. Three key findings were observed: Firstly, as predicted, adults showed greater selectivity for social interactions than children in the pSTS; this effect was strong in the left hemisphere, where children showed no selectivity, but only marginal in the right hemisphere, where children showed relatively strong interaction selectivity. Secondly, adults demonstrated strong, focally tuned interaction responses in the bilateral pSTS, whereas children show weaker, more diffuse responses in the right hemisphere only. Thirdly, unlike children, adults showed

additional interaction selective coactivations in other 'socially tuned' temporal lobe regions (as well as the precuneus). Together, these findings demonstrate that neural responses to social interactions are not fully mature in preadolescent children, and therefore must undergo substantial development during adolescence.

Age Group Differences

Both adults and children showed significantly greater interaction selectivity in the right pSTS than all other ROIs, and this was supported by the whole brain findings that show this was the most active region. These findings are consistent with previous accounts that the pSTS – especially in the right hemisphere – is strongly responsive to dyadic social interactions (Georgescu et al., 2014; Hafri, Trueswell, & Epstein, 2017; Iacoboni et al., 2004; Isik et al., 2017; Kujala, Carlson, & Hari, 2012a; Lahnakoski et al., 2012; Walbrin et al., 2018).

The main prediction of this study was that adults would show stronger interaction selective responses in the pSTS than children, in line with analogous trends for functionally localized face (Scherf et al., 2007) and body regions in the STS (Ross et al., 2014). These findings may be interpreted as children showing weaker, underdeveloped responses in category specific regions. Although this effect was strong in the left pSTS, it was only marginal in the right pSTS, with children showing weaker but comparable responses to adults in this region. Additionally, adults showed strong responses in both hemispheres, and greater responses in right than left pSTS. By contrast, children were selective in the right pSTS only, demonstrating differential activation for bilateral pSTS between age groups.

Interestingly, similar trends have been observed before. Bonte et al. (2013) calculated laterality scores for voice selectivity in the STS (i.e. the ratio between the magnitude and extent of left and right lateralized STS responses) and found that both children and adults showed rightward lateralization, but crucially, this effect was significantly stronger in children, as proportionally greater recruitment of left STS was shown for adults. This study also showed another parallel with the current findings; voice selective responses in the STS were more diffuse and less selective in children than adults (who showed strongly selective and spatially constrained selectivity), mirroring age related differences in focal tuning of interaction selectivity in the current study. Between

these results, a more complex characterization of neural maturity of social responses in the STS is implicated; that is, selective responses become more bilateral and focally tuned across development.

A further developmental difference was observed between age groups in the current study – unlike children, adults showed complementary interaction selective responses in regions neighbouring interaction selective pSTS (e.g. face selective STS cortex and to a marginal extent, left EBA). Along with strong focal tuning in the pSTS, weaker selectivity in neighbouring face selective STS is somewhat unsurprising, given the relevance of facial information in interactive contexts. Indeed, a similar pattern was shown for face information, whereby stronger face selective responses were found in face selective STS, with weaker selectivity in neighbouring interaction selective pSTS in adults (see chapter 4 appendix G & H, for face and body selectivity analyses, respectively). These results suggest dissociable selectivity between these regions, but with accompanying weaker selectivity in each respective 'non-selective' region (e.g. weaker interaction selectivity in face STS). It is also worth noting that PSC for the interaction condition in the face selective STS was substantially weaker than in interaction selective pSTS, further emphasizing the strong response for interaction information in this region.

By contrast, bilateral TPJ was not sensitive to interaction information (i.e. PSC change was not above-zero in TPJ regions, for either adults or children). However, PSC responses did tend to be higher for the interaction than non-interaction condition. This is somewhat consistent with prior univariate evidence that shows no difference between these two conditions with the same stimuli (Isik et al., 2017) but a small difference between interacting and non-interacting moving shapes (Walbrin et al., 2018). These slightly divergent results may reflect stimulus driven differences between studies (e.g. abstract moving shape stimuli may recruit greater mentalizing oriented TPJ resources that are not required to understand relatively more familiar body stimuli).

Left pSTS Responses to Interactions in Children

It is also worth noting that 'non-selectivity' for interactions in the left pSTS in children is not the result of weaker interaction responses per se; instead, strong PSC responses were found for both interactions and non-interactions, but they were approximately equal in magnitude. This could be interpreted in two ways. Firstly, that

responses in this region are driven by the mere presence of two individuals, but not by interactive information (e.g. contingent actions and facing direction). As such, immature responses in this region may reflect simplistic 'baseline' representations of interactions as merely two people together, that are insensitive to nuanced dynamic information that adults make use of to distinguish these two conditions. And secondly, responses in this region may simply reflect sensitivity to biological motion per se, that was essentially equivalent between interaction and non-interaction stimuli. This is supported by previous evidence of STS sensitivity to biological motion in children (Mosconi et al., 2005; Carter & Pelphrey, 2006) and additionally, PSC responses to bodies (from the dynamic localizer task) were comparable to interactions in the left pSTS but not right pSTS (see chapter 4 appendix E). Therefore, it seems that left pSTS responses to interactions are specific to biological motion information in children, and become more sensitive to interactive information across development.

Comparison with Similar Findings

The currents findings show a degree of overlap with the only other currently known study to investigate age-related neural changes in interaction perception. Sapey-Triomphe et al. (2017) employed a whole brain analysis with the same contrast (i.e. interaction > non-interaction) with a highly similar stimulus set to the current study and observed strong pSTS responses across adults, adolescents, and children (8-11 years), along with notable activation in pMTG, pITG, MOG, and IFG. These results are partially replicated by the present analyses, which found strong right pSTS along with aSTS activation in both groups, but pMTG, pITG, and minor IFG responses were only shown for adults.

pSTS and aSTS responses have been shown previously shown for social interactions (Lahnakoski et al., 2012), and dynamic faces (Pitcher et al., 2011), and may suggest functional coupling between these regions, although whether these regions perform similar or different computations during interaction perception remains to be seen. Despite some overlap with the findings of Sapey-Triomphe et al. (2017) differences with the current findings may have occurred for two main reasons: Firstly, the use of an explicit judgement task by Sapey-Triomphe et al. (2017) may have driven mentalizing or attentionally enhanced activations across the brain; and secondly, the inclusion of

younger children in the current study (with presumably weaker responses than older children) may have reduced averaged responses across the child group.

Both studies also did not show pSTS differences for the adults > children whole brain contrast (interaction > non-interaction), suggesting that the nuanced findings observed in the present ROI analysis are not detectable with less sensitive whole brain analysis; this may in part be because these ROI analyses were relatively unaffected by inter-subject variability in the location of pSTS activations, whereas whole brain analyses are. Similarly, this variability may also account for the absence of left pSTS activation for adults in the whole brain analysis (as well as for the adults > children comparison); indeed, significantly greater inter-subject variability in the morphology of the left pSTS compared to right pSTS has been previously observed (Bonte et al., 2013).

Additionally, adults but not children, also showed responses in the precuneus, and this was also one of several regions to show increased activation with age by Sapey-Triomphe et al. (2017). Although the precuneus is typically associated with mentalizing processes (e.g. Schurz, Radua, Aichhorn, Richlan, & Perner, 2014), the absence of mPFC activation in the current study suggests that this is not the case. Instead, precuneus activation has been demonstrated to differentiate between congruent and incongruent dynamic social interactions (i.e. interactors actions from the same corresponding scenario, or different scenarios) while subjects made (non-social) judgements (Petrini, Piwek, Crabbe, Pollick, & Garrod, 2014). Although congruence was not explicitly manipulated in the current study, it is speculated that the qualitatively different congruence information between interaction and non-interaction stimuli may account for the precuneus responses (e.g. interactions depicted two congruent actions, whereas the non-interaction conditions did not). Alternatively, the precuneus, as well as aSTS, show fine-grained sensitivity to the direction of others' eye gaze (Carlin, Calder, Kriegeskorte, Nili, & Rowe, 2011); it could be speculated that the presently observed recruitment of these regions in adults corresponds to the differential facing direction information present in interactions (facing towards each other) and non-interactions (facing away from each other).

Strengths & Limitations

The current findings are supported by several methodological strengths. The currently observed group differences are demonstrably independent of head motion; the use of relatively stringent head motion thresholds, equivalent head motion for each group (for condition pairs used to calculate selectivity measures), and highly comparable PSC values between groups show that group selectivity differences are not an artefact of differential head motion. In addition to demonstrating evidence consistent with focal tuning differences in the pSTS to social interactions, this analysis revealed another important characteristic of the data; that is, that group differences in interaction selectivity of the pSTS were not specific to the ROI size used in the main analyses. Instead, this trend was consistent across all ROI sizes, although this difference was reduced at larger ROI sizes.

It is also worth briefly addressing several limitations of the current study. Firstly, although the study aimed to compare adult responses to pre-adolescent children, a relatively broad age range for children was used, and it is possible that older children demonstrate relatively more adult-like responses than the younger children. Secondly, although responses to interactions and faces are relatively dissociable in the STS, some overlap in these responses was observed. Although the interaction stimuli contained no face information, it is possible that the face stimuli used here contained interactive signals (e.g. facial expressions and orienting movements that might imply the presence of another person or interactor). Thirdly, no formal power analysis was calculated to determine the sample sizes, and so the current group comparisons are not adequately powered (i.e. to detect small effects at 80% power); instead, group sizes were determined via opportunity sampling (i.e. the availability of children aged between 6 – 12 years determined the size of both adult and child age groups). The current group sizes fall short of recommended group sizes for typical univariate fMRI effects (i.e. in the order of 40 – 50 subjects per group; Mumford, 2012; Yarkoni, 2009), and therefore these effects must be interpreted with caution.

One further potential limitation concerns the use of mentalizing localizer task; firstly, the mentalizing > pain contrast may have captured additional attentional or top-down differences between contrasted conditions (e.g. mentalizing likely commands greater top-down attention than pain); and secondly, definition of TPJ with this contrast

may have resulted in differential localization than with other mentalizing localizers, for example when contrasting responses to written false-belief stimuli with non-false-belief stimuli (Dodell-Feder et al., 2011). However, this is unlikely, as direct comparisons between these two tasks demonstrate that they both recruit highly similar regions of TPJ cortex (Jacoby, Bruneau, Koster-Hale, & Saxe, 2016). Potential concerns over the specific contrast used to localize TPJ are further alleviated by the specific aim of this study; that is, to test interaction responses in pSTS (and other regions, including TPJ), rather than to test mentalizing responses per se. Further to this, if additional top-down attentional differences do contribute strongly to responses in the TPJ, such attentional differences do not contribute strongly to the interaction task responses; PSC responses were not greater than zero for any of the three interaction conditions in this TPJ region, for either age group (see chapter 4 appendix E).'

Conclusions

In summary, the present findings demonstrate that pre-adolescent children show markedly different neural responses to dyadic social interactions than adults. These results imply that the maturation of neural interaction selectivity is characterized by increasingly bilateral pSTS activation that exhibits strong focal tuning, along with weaker responses in neighbouring STS cortex. These findings should motivate further research to explore interaction responses across adolescence and into late adulthood.

Chapter 5

Dyadic Interaction Processing in the Posterior Temporal Cortex

Abstract

Recent behavioural evidence suggests that individuals engaged in social interactions are not simply encoded as separate individuals, but are perceived and memorized as an integrated unit, or 'more than the sum of their parts'. However, it is unknown exactly where in the brain such integration occurs. Converging functional magnetic resonance imaging (fMRI) evidence demonstrates the important role that posterior temporal cortex – especially the posterior superior temporal sulcus (pSTS) – plays in visually perceiving social interactions. The current study aimed to investigate whether the pSTS or other posterior temporal lobe regions: 1) demonstrate evidence of a *dyadic information effect* – that is, qualitatively different responses to an interacting dyad than to averaged responses of the same two interactors, presented in isolation, and; 2) significantly differentiate between different types of social interactions.

Multivoxel pattern analysis was performed in which a support vector machine classifier was trained to differentiate between qualitatively different types of dyadic interactions. Above-chance classification of interactions was observed in 'interaction selective' pSTS-I and extrastriate body area (EBA), but not in other regions of interest (i.e. face selective STS and temporo-parietal junction). A dyadic information effect was not observed the pSTS-I, but instead this trend was shown in the EBA; that is, classification of dyadic interactions did not fully generalise to *averaged responses* of the isolated interactors, indicating that dyadic interaction representations in the EBA contain unique information that cannot be recovered from the isolated interactors. Follow-up analyses suggest that the respective 'action-gestures' of the two interactors contribute strongly to interaction representations in this area. These findings suggest that the EBA might play an important role in the integration of interacting individuals into a unified interactive representation.

5.1. Introduction

Recent behavioural evidence suggests that visual displays of two interacting individuals are not merely processed as separate individuals, but instead are perceived (Papeo et al., 2017; Vestner, Tipper, Hartley, Over, & Rueschemeyer, 2018) and remembered (Ding, Gao, & Shen, 2017) as an integrated unit that is 'more than the sum of their parts'. But where in the brain might this dyadic integration occur? Recent fMRI research demonstrates the importance of the posterior superior temporal sulcus (pSTS) for processing dyadic social interactions (Isik, Koldewyn, Beeler, & Kanwisher, 2017; Walbrin, Downing, & Koldewyn, 2018), and similarly, regions of the STS are implicated within a wider brain network when viewing social interactions in the rhesus macaques (Sliwa & Freiwald, 2017). However, whether this region plays a key role in integrating the actions of interactors is currently unknown.

Integrative processing of dynamic social information in the STS has been demonstrated previously. For example, Calvert, Campbell, and Brammer (2000) found 'supra-additive' responses to congruent audio-visual stimuli (i.e. responses to auditory voice stimuli and video footage of corresponding mouth movements are significantly greater than responses to each modality, summed together); crucially, such responses suggest that multimodal representations contain information that is qualitatively different than responses from each modality presented in isolation. Similar audio-visual integration in the STS is also shown for visually presented objects and their corresponding sounds (e.g. seeing a phone and hearing it ring; Beauchamp, Lee, Argall, & Martin, 2004), and the STS is implicated as the locus of perceptual illusions that arise from audio-visual discrepancies (Nath & Beauchamp, 2012).

By contrast, configural integration – of multiple *separate* objects – is observed in object selective temporal lobe cortex that lies ventrally to the STS. Voxel pattern responses in lateral occipito-temporal cortex (LOTC) are qualitatively different, depending on the spatial configurations of two objects; 'regular' spatial configurations of objects (e.g. a sofa positioned in front of a television) are *less similar* than 'irregular' configurations (e.g. a television positioned behind a sofa) are to the average response to isolated objects (e.g. sofa and television presented separately; Kaiser & Peelen, 2018). These findings suggest that neural representations of objects are not merely sensitive to the presence of objects, but also encode information about the spatial relations between

objects. Similarly, stronger univariate LOTC responses are evoked by images of two objects that are positioned to imply an action (e.g. a pitcher positioned above and tilted towards an empty glass) than when not arranged to imply an action (Roberts & Humphreys, 2010).

Indeed, similar responses are shown for human-object interactions in object selective LOTC, and interestingly, in pSTS (Baldassano, Beck, & Fei-Fei, 2017). Voxel pattern responses to images of either human-object interactions (e.g. a human pushing a cart, or typing at a computer), or isolated humans and isolated objects (taken from the same human-object interaction images) were compared to determine which brain regions represented these interactions as 'more than the sum of their parts'. It was found that responses to human-object interactions did not fully generalise to 'pattern averages' created from isolated humans and objects in pSTS and object-selective LOTC; it was concluded that representations in these regions are specifically sensitive to interactive information that could not be accessed from the isolated 'parts' of these interactions.

The present functional magnetic resonance imaging (fMRI) study aimed to investigate emergent responses to dynamic social interactions between two individuals. It was hypothesized that the pSTS would show a *dyadic information effect* – that is, that responses to dyadic interaction stimuli would not fully generalise to averaged responses evoked by each interactor presented in isolation. This prediction was motivated by several previous observations of pSTS responses: Strong sensitivity to dynamic social interactions (Isik et al., 2017; Walbrin et al., 2018); non-linear, configural processing of static human-object interactions (Baldassano et al., 2017), and; integration of *dynamic* multi-modal social information (Beauchamp et al., 2004; Nath & Beauchamp, 2012; Calvert et al., 2000). Additionally, it was predicted that significantly differentiable responses to different types of interaction would be observed in the pSTS, replicating previous findings (e.g. Isik et al., 2017; Walbrin et al., 2018). Responses were also tested in 3 other posterior temporal lobe regions of interest (ROIs) that might plausibly also show this effect (i.e. these regions are all selective for different categories of social information and may be strongly recruited when viewing social interactions): Extrastriate body area (EBA), temporo-parietal junction (TPJ-M), and face selective STS (STS-F).

5.2. Methods

Participants

21 adults (mean age = 23.40 years; SD = 3.74; range = 18 – 35; 12 females) were scanned in two separate sessions – one for the main task, and another session on a separate day for the localizer tasks. Hand preference was determined using the Edinburgh Handedness Inventory (Oldfield, 1971) to ensure that all subjects were right hand dominant (due to increased functional lateralization differences observed in left-handed individuals). Participants gave informed consent and received monetary compensation for taking part. Ethical procedures were approved by the Bangor University ethics board.

Stimuli

Stimuli consisted of 4 second (s) video clips that were taken from custom footage of two actors engaging in semi-improvised interactions; actors were instructed to improvise these scenarios while enacting scripted body-based actions or '*action-gestures*' that they were encouraged to perform in a natural, authentic way (see chapter 5 appendix A for a full description of stimulus creation and scripted gestures). An initial set of *dyad stimuli* (see figure 5.1) depicted two interactors engaging in one of three interactive scenarios: *Arguing* (i.e. both actors engaged in an angry/frustrated confrontation), *celebrating* (i.e. both actors celebrated together, excitedly), and *laughing* (i.e. both actors were laughing together, or at each other in a playful manner). These specific scenarios were chosen for the 'tonal consistency' of actions performed by a given pair of interactors, such that the intentions, emotions, and valence information conveyed by both individuals in a given scenario were always similar (e.g. angry/frustrated) rather than contrasting (e.g. angry/sad). This ensured that successful classification of the different scenarios was not driven by systematic differences in intentional, emotional, or valence content *between interactors*. However, systematic differences in this information across the three interactive scenarios likely did contribute to classification (e.g. negative and positive valence for arguing and celebrating, respectively). As such, these scenarios represented three highly distinct interactive scenarios that were intended to be easily distinguishable.

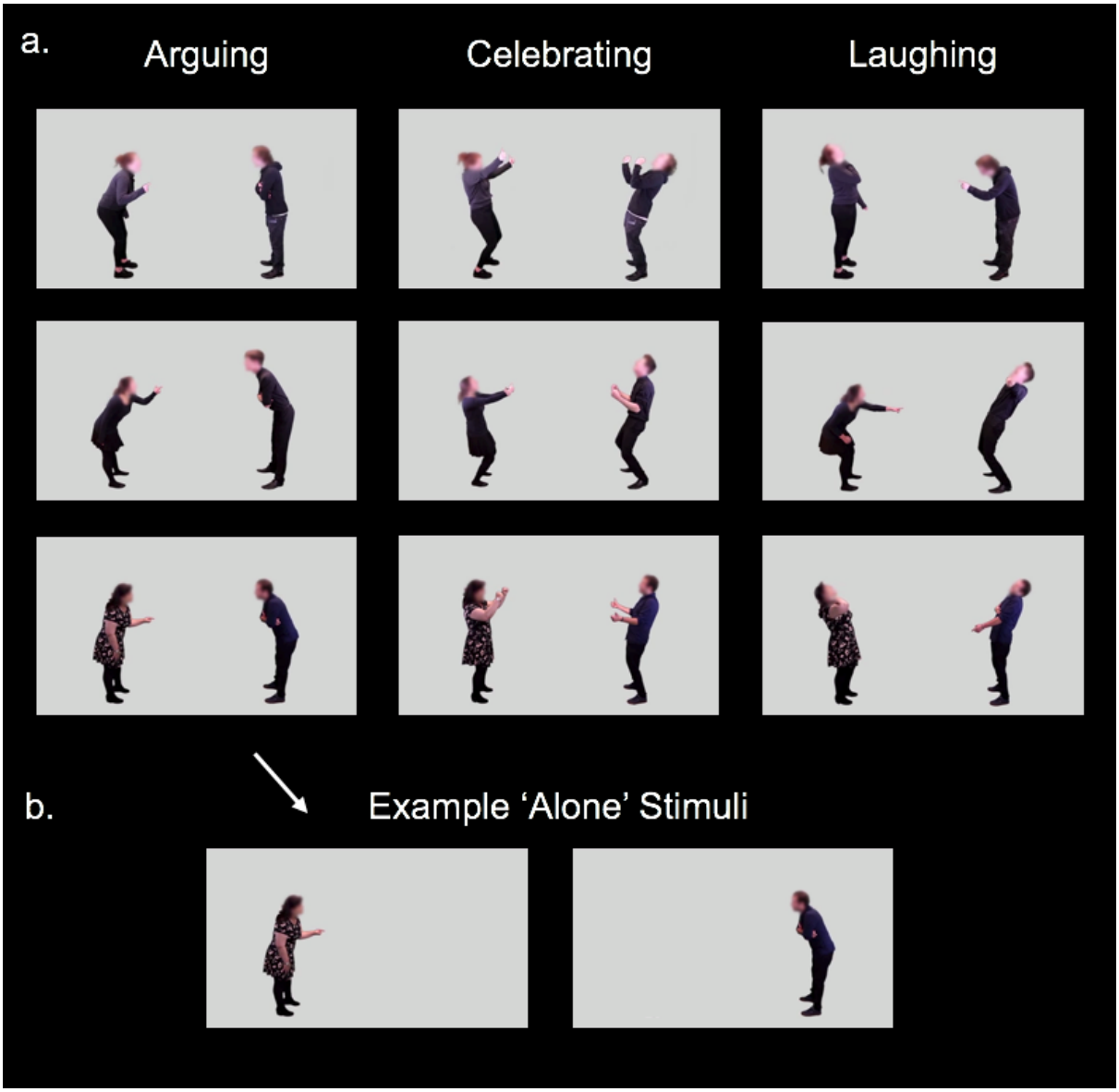


Figure 5.1. a. Example frames from the three dyad condition video stimuli. Each row represents one of three unique female-male interactor pairs performing examples of each of the three interaction scenarios. b. An example pair of alone stimuli in which a single interactor was presented in isolation.

Within each interaction scenario, four unique *action-gesture* pairings were created. Specifically, eight unique action-gestures were depicted *once* across four separate videos, such that each video showed the two individuals performing a complementary pair of action-gestures (e.g. while arguing, interactor A angrily points in an accusatory manner at interactor B who is shaking their hands in frustration). Importantly, no gestures were 'reused' in any of the other action-gesture pairings (see chapter 5 appendix A for descriptions of each action-gesture). Similarly, 3 different female-male *interactor pairs* enacted these scenarios, yielding a total of 36 dyad stimuli: 3 interaction scenarios (*arguing, celebrating, laughing*) x 4 unique action-gesture pairings

x 3 interactor pairs. These stimuli were chosen from a wider set of stimuli based on the highest 'interactiveness' and 'naturalness' ratings from a pilot study (N=10; see chapter 5 appendix A for further details).

For these stimuli, the average horizontal distance between actors was closely matched (i.e. visual angle between the centre of each actor's torso was approximately 4.80°), and actor height ranged between 3.73 – 4.26°. As dynamic facial information is known to drive responses to dynamic facial information in the STS (e.g. Deen, Koldewyn, Kanwisher, & Saxe, 2015), the presence of facial information was controlled such that classification could not be attributed to different facial expressions. Accordingly, these stimuli did not contain high spatial frequency face information, but body information was preserved. To achieve this, a circle shaped Gaussian blur mask was placed on each of the actors' heads for each video frame. This preserved the overall shape of the head, preventing the potentially eerie appearance of headless interacting bodies.

To test neural responses to the same interactive information – but without specifically *dyadic information* (e.g. without temporally correlated movements between the two individuals) – a separate set of 72 *alone stimuli* were created by removing either individual from each of the 36 dyad stimuli. It is important to note that, although these stimuli depicted an isolated interactor by themselves, they still conveyed interactive information (e.g. communicative gesturing towards an implied interactor). Two horizontally 'flipped' variants of these 108 unique stimuli (36 dyad + 72 alone stimuli) resulted in a final set of 216 stimuli.

Design & Procedure

A rapid event-related design was used, and each run was optimized using optseq2 (http://surfer.nmr.mgh.harvard.edu/optseq), based on differentiating six conditions (i.e. both dyad and alone variants of the arguing, celebrating, and laughing interaction scenarios), with an inter-stimulus interval range between 0s – 10s (along with 8s fixation at the beginning of each run, and 16s at the end to capture most of the haemodynamic response). The six designs with the highest detection sensitivity were selected to configure event timings for runs.

In the scanner, participants viewed stimuli that were presented centrally on the screen within a 9.17 x 5.11° rectangular space. Six runs were completed, each lasted

exactly 7 minutes and contained 8 and 16 stimuli for each dyad and alone version of the three interaction scenarios, respectively, resulting in 72 stimuli per run. Two important stimulus ordering considerations are also noted here: Firstly, that any given pair of alone stimuli (i.e. that originated from the same dyad stimulus) were always presented in the same run as each other so that classification of alone stimuli did not contain additional between-run variance that was not present for the dyad stimuli; secondly, to minimize repetition effects (i.e. seeing the exact same action-gestures from a given dyad stimulus and the corresponding pair of alone stimuli), alone stimuli that appeared in any given run were always from dyad stimuli that were allocated to a different run.

In addition to the stimuli already described, nine additional catch stimuli were presented (three dyad stimuli, and six alone stimuli) but were not later analysed. These trials contained a 'frame freeze' in which 12 consecutive video frames (duration = 500ms) were randomly removed from the video and replaced with one repeated frame for that period, creating the impression of a momentary video pause. Participants were instructed to simply watch the videos and to give a button-press response whenever a frame freeze was detected. Additionally, participants were told to not think about or make explicit judgements about the scenarios depicted in the videos, so as to minimize the influence of explicit mentalizing (e.g. 'actor A is angry with actor B').

Localizer Tasks and ROI Creation

Participants completed several localizer tasks in a separate scanning session, on a separate day (these localizer tasks are described in detail in chapter 4). Briefly explained, 4 bilateral ROIs were localized with the corresponding tasks: pSTS-I (interaction selective pSTS: point-light interactions > scrambled interactions); STS-F (face selective STS: dynamic faces > objects); EBA (dynamic bodies > objects); and TPJ-M (Pixar video mentalizing localizer: mentalizing > pain). A group-constrained ROI definition procedure (e.g. Julian, Fedorenko, Webster, & Kanwisher, 2012) was used to create subject-specific ROIs, as follows. For a given subject, and contrast (e.g. interaction > scrambled interaction, for the pSTS-I), a 5mm-radius 'search sphere' was created by running a whole brain analysis for N-1 group subjects (i.e. with the current subject excluded) and centring the sphere at the peak voxel (i.e. highest t-value) in the designated region. This relatively small sphere was used to ensure subject's ROIs did not deviate too far from a given

designated anatomical region (e.g. pSTS). To determine the position of the final ROI, a whole brain analysis for the current subject (for the same contrast) was run, but resulting activation was constrained to the search sphere. A 7mm-radius sphere was then centred at the peak voxel in this constrained region. This sphere size was chosen as an ideal compromise between capturing a relatively large number of voxels that would allow stronger classification (e.g. Coutanche, Solomon, & Thompson-Schill, 2016), and ensuring minimal overlap between neighbouring ROIs.

All ROIs contained 179 voxels, with the exception of two subjects that had small regions of overlap between the right pSTS-I and right TPJ-M, and a further two subjects with similar overlap between the right pSTS-I and right STS-F. Across these four subjects, a mean overlap of 18 voxels (range: 12-24) was found. To ensure independence of ROI voxels within each of these four subjects, overlapping voxels were removed and ROIs were recreated (respective final ROI sizes for these four subjects were: 167, 161, 161, 155 voxels; all other ROIs for these subjects contained 179 voxels).

MRI Parameters

Scanning was performed with a Philips 3T scanner at Bangor University. Functional images for the main task were acquired with the following parameters: a T2*-weighted gradient-echo single-shot EPI pulse sequence; TR = 2000ms, TE = 30ms, flip angle = 83°, FOV(mm) = 240 x 240 x 108, acquisition matrix = 80 x 78 (reconstruction matrix = 80); 36 contiguous axial slices were acquired, and reconstructed voxel size was $3mm^3$. Four dummy scans were discarded prior to image acquisition for each run. Functional images for the localizer data differed slightly from those used for the main task, as this data was also used for a developmental study with children. These parameters only differed from the main task as follows: FOV(mm) = 240 x 240 x 112, 32 contiguous axial slices in ascending order, acquired voxel size (mm) = 3 x 3 x 3.5 (reconstructed voxel size = $3mm^3$). Structural images (for both main task session and localizer session) were obtained with the following parameters: T1-weighted image acquisition using a gradient echo, multi-shot turbo field echo pulse sequence, with a five echo average; TR = 12ms, average TE = 3.4ms, in 1.7ms steps, total acquisition time = 136s, FA = 8°, FOV = 240 x 240, acquisition matrix = 240 x 224 (reconstruction matrix =

240); 128 contiguous axial slices, acquired voxel size(mm) = 1.0 x 1.07 x 2.0 (reconstructed voxel size = $1mm^3$).

fMRI Pre-processing and GLM Estimation

Pre-processing was performed in SPM12 (fil.ion.ucl.ac.uk/spm/software/spm12) for each data set separately (i.e. data from the main task session, and data from the localizer tasks session), with the following steps: Realignment (and reslicing), co-registration, segmentation, normalization, and smoothing. All default parameters were used except for a 6mm FWHM Gaussian smoothing kernel. Additionally, slice-timing correction was performed prior to realignment for the event-related data from the main task, but not for the blocked design localizer data as minimal estimation cost has been previously shown for blocks >15s (Sladky et al., 2011).

General linear model (GLM) analysis was performed in SPM12 on participants' normalized images (for both smoothed and unsmoothed versions) for both main task and localizer datasets, separately. Event durations and onsets for each run were modelled using a boxcar reference vector and convolved with a canonical hemodynamic response function (without time or dispersion derivatives), with a high-pass filter of 128s and autoregressive AR(1) model. Rest periods were implicitly modelled and the 6 motion parameters were modelled as nuisance regressors. Catch trials were also modelled separately, but not analysed.

Two sets of whole brain beta maps were generated on a run-wise basis, for different purposes. For classification analyses, events were modelled as 6 *classification conditions* – both dyad and alone variants of the arguing, celebrating, and laughing stimuli. Therefore, each run was modelled as 6 beta maps (along with intercept, motion parameters, and catch trial condition images). For stimulus-wise analyses, individual events from each run were modelled – that is, 72 beta maps were generated for the 72 stimuli shown in a given run (along with intercept, motion parameters, and catch trial images).

SVM Classification Analyses

Leave-one-run out classification was implemented with CoSMoMVPA (Oosterhof, Connolly, & Haxby, 2016) with a linear support vector machine (SVM) approach. Briefly explained, for a given subject, an SVM classifier was trained on ROI voxels (i.e. beta values) for the conditions of interest (e.g. dyad variants of the arguing, celebrating, and laughing conditions) in all but one run of data, with the 'left-out' run of data used to independently test classification performance on. This was iterated six times with each run serving as the left-out test run, and classification accuracy was averaged across iterations. These values were then entered into group level t-tests. All tests are significant at the corrected Bonferroni threshold (α) unless otherwise stated. A corrected Bonferroni α was calculated separately for each set of analyses, as stated in each sub-section in the results. All t-test *p*-values are one-tailed.

This approach was almost identical for both 'standard' classification (e.g. between the three dyad conditions, or between the three alone conditions) and cross-classification analyses except that training and test conditions differed; for example, the classifier was trained on the three dyad conditions, but *tested* on the three alone conditions. Significant cross-classification demonstrates that the patterns underlying the two sets of conditions are qualitatively similar to each other, and therefore are largely driven by the same information. Crucially, in the present study, it was reasoned that if a region showed significantly greater dyad classification than cross-classification (i.e. between dyad and alone conditions), this would indicate sensitivity to dyadic information that could not be 'recovered' from the constituent halves of these dyads presented in isolation (i.e. averaged responses to alone stimuli).

For searchlight analyses (Kriegeskorte, Goebel, & Bandettini, 2006), an identical SVM classification approach to that described above was adopted, with the following exception: Rather than calculating a single classification accuracy value for an ROI, a 'searchlight' sphere (radius = 5 voxels) was iteratively moved across the brain (i.e. the sphere was centred on each voxel) and pattern classification was performed on all searchlight voxels, with the average classification value assigned to the central voxel in an output image. This resulted in a whole brain 'searchlight map' of classification accuracy values for each participant.

For group level inference of searchlight data, a threshold-free cluster enhancement (TFCE; Smith & Nichols, 2009) approach was implemented; unlike other forms of cluster-wise whole brain inference that typically rely on the relatively arbitrary setting of an activation height threshold (i.e. minimal voxel activation threshold), TFCE typically offers a more sensitive means of cluster level inference. Specifically, clusters are defined by the magnitude and extent of activation in neighbouring voxels and then inference is made by comparisons against a null distribution. Accordingly, the TFCE procedure was applied to each subject's searchlight map(s) and then compared against 10'000 Monte Carlo permutations. The resulting group level z-score maps describe the degree of above-chance classification and 'local cluster support' of voxels.

Representational Similarity Analyses

Representational similarity analysis (RSA; Kriegeskorte, Mur, & Bandettini, 2008) was used to compare voxel pattern responses to individual dyad stimuli with 'artificial pairs' of alone stimuli in which the *average information* differed from the corresponding dyad stimuli; for example, for a given pair of stimuli, one half exactly matched half of the corresponding dyad stimulus, but the other varied on a given dimension, such as the identity of the interactor. This artificial pairs analysis was applied on a per-subject, per-region basis, and was implemented as follows (see figure 5.2). A dyad RDM was created, such that each cell represented the 'dissimilarity' (i.e. 1 – Spearman's *r* coefficient) between voxel responses for each given pair of dyad stimuli. Given the 36 unique dyad stimuli, a 36 x 36 matrix was generated that depicted a total of 630 unique comparisons (i.e. after removing 36 'self-correlation' and 630 duplicate values: ((36*36)-36)-630). It is worth noting that each of the 36 stimulus patterns corresponded to the mean pattern of all presentations for that stimulus across the six runs.

Next, a series of artificial pair RDMs were created in a similar manner, but with one crucial distinction; before constructing these RDMs, voxel patterns corresponding to *specific pairings* of 72 alone stimuli were averaged together (e.g. mean voxel patterns of two given alone stimuli were averaged together to create a single mean pattern). These 36 artificial patterns were then used to create an RDM in the same way as described above. Importantly, the ordering of patterns was consistent across RDMs; for example, if the first row of the dyad RDM depicted a specific stimulus (e.g. 'Arguing_Pair1_

ActionGesture1'), then the first row of each artificial pair RDM always corresponded to a pairing containing (at least) one of the individuals from the same dyad stimulus.

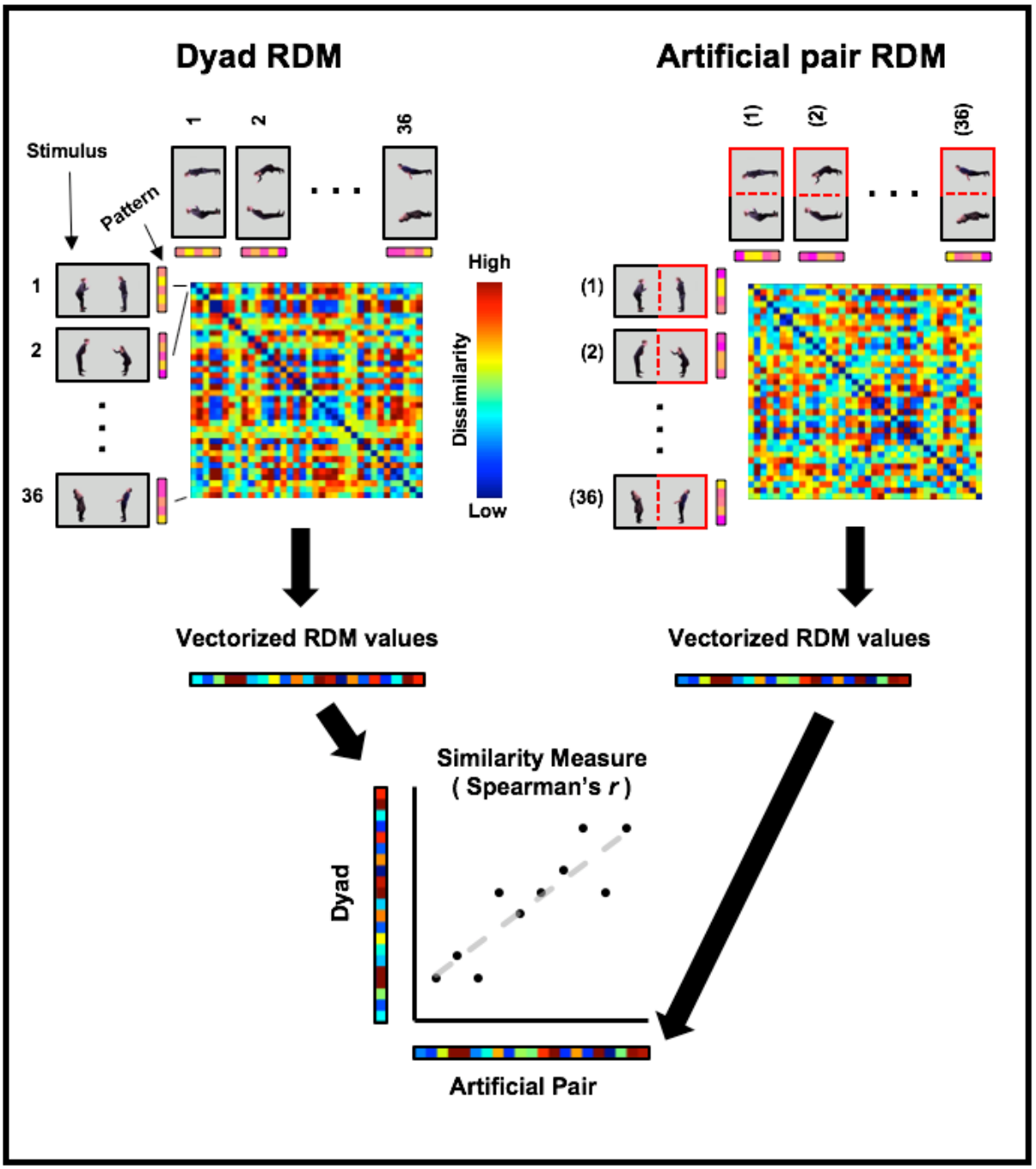


Figure 5.2. An illustration of the artificial pairs analysis. An example dyad RDM is depicted in the top left corner where cells represent the pair-wise dissimilarity (1 – *r*) between patterns evoked by the 36 dyad stimuli. An example artificial pair RDM is shown in the top right corner; here cells represent dissimilarities between *mean patterns* for 36 artificially paired stimuli. RDMs were then vectorized and correlated to calculate a final measure of similarity.

In total, four separate artificial pair DSMs were created, based on separate pairing rules (see section 5.3). The final representational similarity measures were calculated for

each comparison between the dyad RDM and a given artificial pair RDM, by vectorizing the 630 unique values in each RDM and calculating the Spearman's *r* correlation coefficient between them. The resulting correlation coefficients were Fisher z-transformed, before performing group level t-tests with these values.

5.3. Results

SVM Classification: ROI Analyses

For each of the 8 functionally localized ROIs, a series of analyses were performed, in which a linear SVM classifier was trained and tested on different variants of the three interaction scenarios (i.e. arguing, celebrating, and laughing). One-sample t-tests were used to determine whether classification accuracy was above-chance (i.e. 100% / 3 categories = 33% chance accuracy; Bonferroni corrected α for dyad classification = .006).

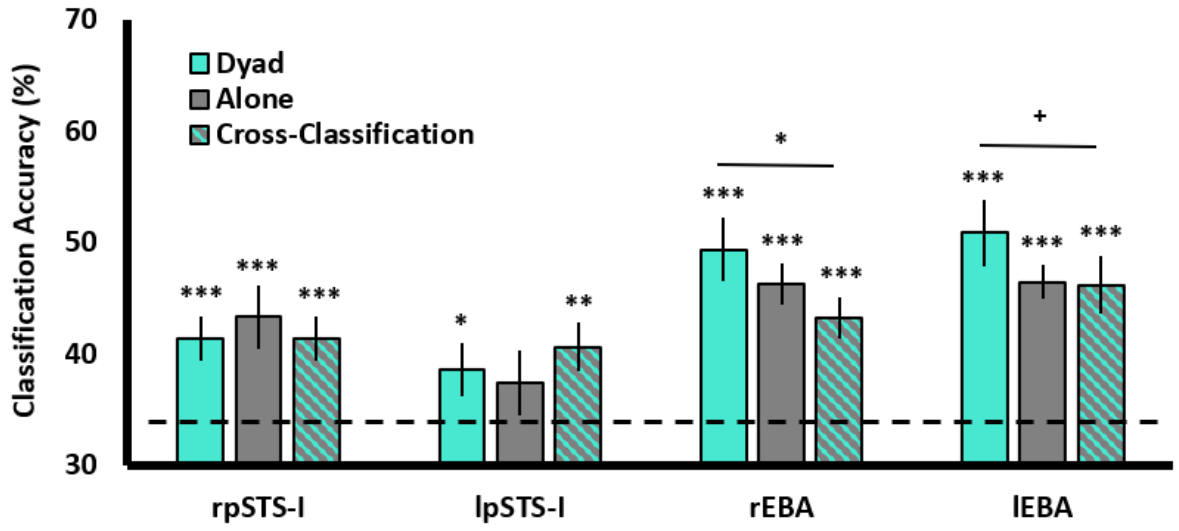


Figure 5.3. A bar chart showing classification accuracy values for dyad, alone, and cross-classification iterations for bilateral pSTS-I and EBA ROIs. Dashed line represents chance-level accuracy (33%). *** = $p \leq .001$; ** = $p \leq .010$; * = $p \leq .05$; $^{+}$ = $p = .073$. Error bars are SEM.

Significant above-chance classification of the three interaction categories of dyad stimuli (see figure 5.3) was observed in the right pSTS-I (Classification accuracy (%): $M = 41.39$, $SD = 9.10$; $t(19) = 3.96$, $p < .001$) and both the right EBA ($M = 49.38$, $SD = 12.19$; $t(17) = 5.59$, $p < .001$) and left EBA ($M = 50.88$, $SD = 13.00$; $t(18) = 5.88$, $p < .001$), and at an uncorrected threshold in the left pSTS-I ($M = 38.60$, $SD = 10.55$; $t(18) = 2.17$, $p = .022$).

None of the four other ROIs – bilateral STS-F and TPJ-M – showed above-chance classification of the dyad stimuli (all *ps* > .100; see figure 5.4; see chapter 5 appendix B for full descriptive statistics).

It is possible that significant classification of dyad stimuli in the bilateral pSTS-I and EBA does not completely rely on inherently dyadic information, and may also encode information conveyed by isolated individuals (e.g. interactive gestures directed towards an implied – but physically absent – interaction partner). To test if this was true, another classification analysis (Bonferroni corrected α = .006) was run with the alone stimuli (see figure 5.3 & figure 5.4). It is worth reiterating that the *same overall information* was present as in the dyad classification analysis (i.e. same scenarios, actors, gestures). Above-chance classification was shown in right pSTS-I (M = 43.33, SD = 12.57; $t(19)$ = 3.56 , p = .001) but only marginally in left pSTS-I (M = 37.43, SD = 12.81; $t(18)$ = 1.39, p = .090). Both right EBA (M = 46.30, SD = 7.86; $t(17)$ = 7.00, p < .001), and left EBA (M = 46.49, SD = 6.73; $t(18)$ = 8.52, p < .001) also showed significant classification. Although bilateral STS-F and TPJ-M regions did not classify dyad stimuli, classification of the alone stimuli was tested in these regions to determine if they may have shown more sensitivity to isolated humans instead. This was not the case, and classification was not above-chance in any of these ROIs (all *ps* > .088). Therefore, these regions were excluded from further analyses.

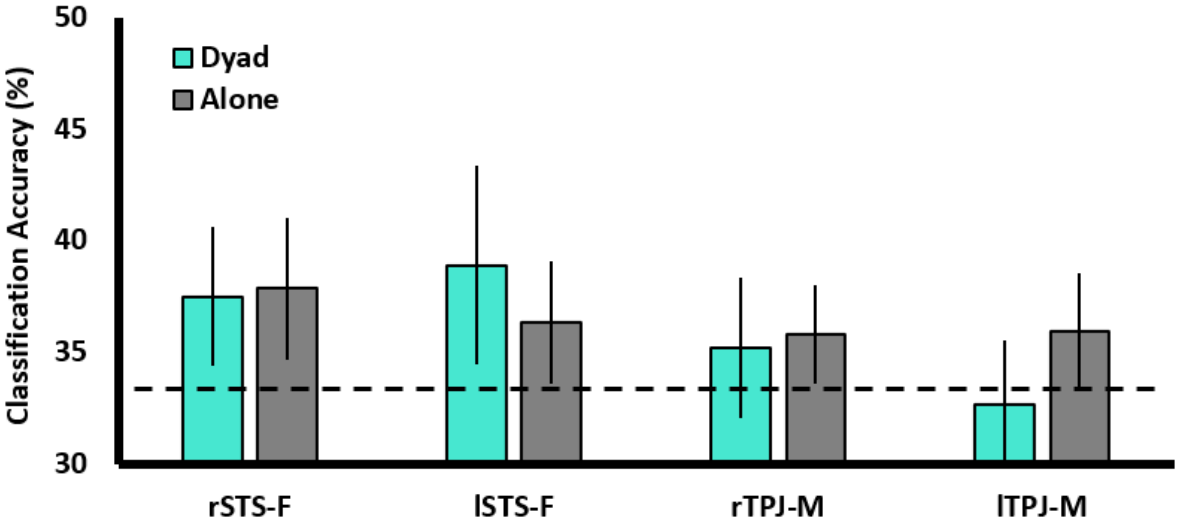


Figure 5.4. A bar chart showing classification accuracy values for dyad and alone classification for bilateral STS-F and TPJ-M ROIs. Dashed line represents chance-level accuracy (33%). No results were significant. Error bars are SEM.

Together, these two classification analyses demonstrate interaction sensitive responses in the right pSTS-I and bilateral EBA regions, and to a marginal extent in the left pSTS-I; specifically, these regions were able to differentiate between the three different interaction scenarios both when observing an intact dyad and when observing the same constituent interactors presented in isolation. However, despite sensitivity to both modes of presentation, this does not mean that the underlying information driving classification in both dyadic and alone scenarios is the same (e.g. information about the spatial-relations between interactors may contribute to classification of the dyad stimuli, but not the alone stimuli). Indeed, if voxel pattern classification in any region does not fully generalise from dyad stimuli to the alone stimuli, this would suggest that there is information encoded by these regions during dyadic interaction perception that cannot be recovered by the same information presented in the alone stimuli.

Next, a cross-classification analysis was implemented (Bonferroni corrected $\alpha = .013$) whereby an SVM classifier was trained to discriminate the three interaction scenarios with the *dyad stimuli*, but was tested on responses to the *alone stimuli*. Significant cross-classification was shown for all four ROIs (right pSTS-I: $M = 41.39$, $SD = 8.92$; $t(19) = 4.04$, $p < .001$; left pSTS-I: $M = 40.64$, $SD = 9.63$; $t(18) = 3.31$, $p = .002$; right EBA: $M = 43.21$, $SD = 7.75$; $t(17) = 5.40$, $p < .001$; left EBA: $M = 46.20$, $SD = 11.27$; $t(18) = 4.97$, $p < .001$), demonstrating that these regions appear to encode similar information in both the dyad and alone stimuli.

To test for the main hypothesis – that is, a dyadic information effect – paired t-tests were then performed (Bonferroni corrected $\alpha = .013$) between dyad classification accuracy scores and cross-classification accuracy scores. No difference was observed for either the right pSTS-I ($t(19) = 0.00$, $p = .500$) or left pSTS-I ($t(18) = -0.73$, $p = .763$), showing no dyadic information effect, indicating that the main hypothesis was not supported. However, significantly greater accuracy for dyad classification than cross-classification was shown in the right EBA ($t(17) = 2.07$, $p = .027$), although this did not survive multiple comparison correction. A similar, although weaker, marginal effect was also shown in the left EBA ($t(18) = 1.52$, $p = .073$). Therefore, evidence suggestive of a dyadic information effect was shown in the bilateral EBA only.

Due to the marginal nature of these results in the EBA, Cohen's *d* effect sizes for both of these tests were calculated. A medium effect size was found for the right EBA ($d = 0.60$), in support of the dyadic information effect. For the left EBA, a small-to-medium

effect was shown (d = 0.38), indicating that this trend was less strongly supported than in the right EBA. To further explore the reliability of these results (i.e. to ensure that these effects were not driven by the arbitrary assignment of classes for cross-classification), cross-classification was performed again, but with the training and testing roles reversed. That is, the classifier was now trained on the *alone* stimuli and tested on the *dyad* stimuli. Both right EBA (M = 43.83, SD = 7.60; $t(17) = 5.86$, $p < .001$) and left EBA (M = 46.20, SD = 10.32; $t(17) = 5.43$, $p < .001$) showed significant cross-classification. Importantly, dyadic information effects were shown to be highly reliable; greater accuracy for dyad classification than cross-classification was again shown in the right EBA ($t(17) = 2.03$, p =.029; d = 0.55) and marginally in the left EBA ($t(18) = 1.41$, p = .088; d = 0.40). Notably, the p-values and effect sizes were very similar to those shown in the initial dyadic information comparisons.

In summary, although right pSTS-I – and marginally, left pSTS-I – differentiated between the three interaction scenarios, no evidence for specific *dyadic information* encoding was observed. Instead, this effect was observed in the right EBA; although this effect did not survive multiple comparison, a medium effect size (and replication with reversed cross-classification) suggests that this is an interpretable effect. Similar trends were also shown in the left EBA (although to a weaker extent), suggesting that these effects are present in both hemispheres.

SVM Classification: Whole Brain Searchlight Analyses

Whole brain SVM classification searchlight analyses were also run to determine whether any regions outside of the EBA showed this dyadic information effect. Whole brain searchlight maps for dyad classification, alone classification, and cross-classification were generated, along with a dyadic information contrast (i.e. dyad classification > cross-classification). For dyad classification (see figure 5.5), strong responses were observed in left LOTC and more focally in the right LOTC, along with bilateral pSTS, superior occipital gyrus, right insula cortex and bilateral early visual cortex. Highly similar responses were shown for both alone (see figure 5.5) and cross-classification analyses (see chapter 5 appendix C). However, no surviving regions were observed for the dyadic information contrast, showing that the nuanced effects observed in the ROI analysis did not survive correction in the whole brain searchlight analysis.

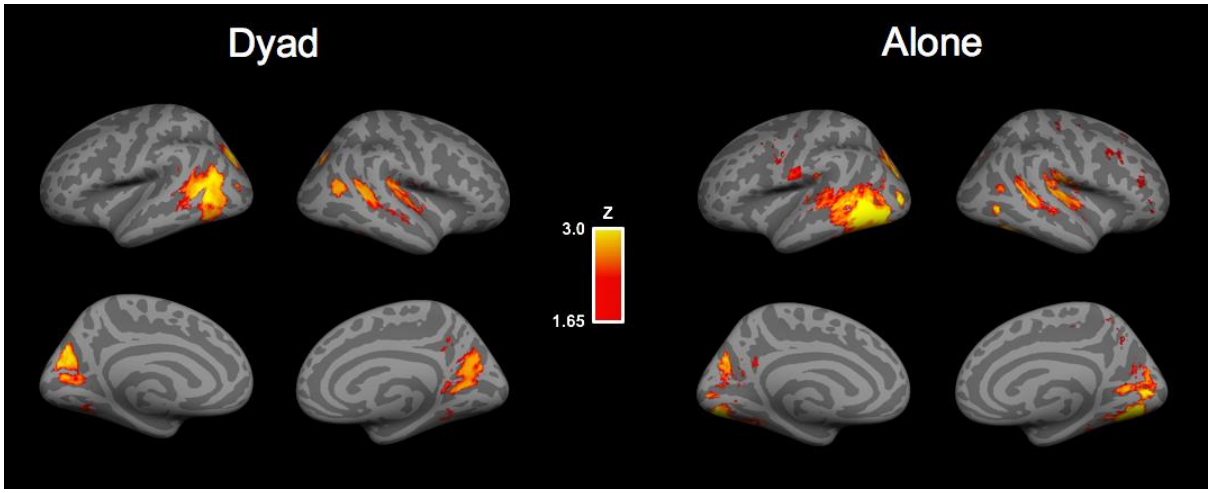


Figure 5.5. Surface-registered whole brain searchlight results for dyad and alone classification. Colour bar represents TFCE z-scores.

Artificial Pairs Analysis

A further exploratory analysis was conducted to examine responses to interaction stimuli in the EBA. It is noted that although this analysis does not directly address which information is driving the dyadic information effect, understanding which information is 'important' to EBA representations of interactions may aid in the interpretation of which information contributes to the dyadic information effect. It is speculated that action-gesture information makes a stronger contribution to EBA representations of interactions than other sources of information in these stimuli (e.g. identity of interactors). This prediction was motivated by prior evidence that the wider LOTC region – in which the EBA is situated – is sensitive to, and able to differentiate between, a wide variety of observed actions (e.g. Hafri, Trueswell, & Epstein, 2017; Wurm, Caramazza, & Lingnau, 2017).

For this RSA approach, RDMs were generated (for each subject, for each ROI) from voxel pattern responses to all 36 dyad stimuli, and correlated with RDMs based on 36 *artificial pairs* of alone stimuli. In total, four separate artificial pair RDMs were created: 1) A *same* RDM (each artificial pair of stimuli exactly matched the corresponding dyad stimulus for all content); 2) An *identity* RDM (the identity of one of the artificially paired interactors was swapped, but action-gesture information was the same as the corresponding dyad stimulus); 3) An *Action* RDM (the action-gesture of one of the artificially paired interactors was swapped for another gesture from the same interaction

scenario, performed by the same individual); 4) *Scenario* RDM (the action-gesture of one of the artificial pairs was swapped for another gesture from a *different* interaction scenario, performed by the same individual).

Similarity was expressed as the correlation between the dyad RDM and a given artificial pair RDM; for example, the *same* similarity measure described the correlation between the *same* RDM and the dyad RDM (similarity measures for action, interaction scenario, and identity conditions were similarly generated by correlating each of these respective RDMs with the dyad RDM). The *same* similarity measure served as a 'baseline' measure (i.e. effectively all information was the same in both dyad and *same* RDMs) against which the other three similarity measures were tested with paired *t*-tests. We reasoned that if action-gesture information contributes strongly to interaction representations in the EBA, significantly less pattern similarity would be found for conditions in which action information is disrupted (i.e. *action* and *scenario,* but not *identity*, similarity measures).

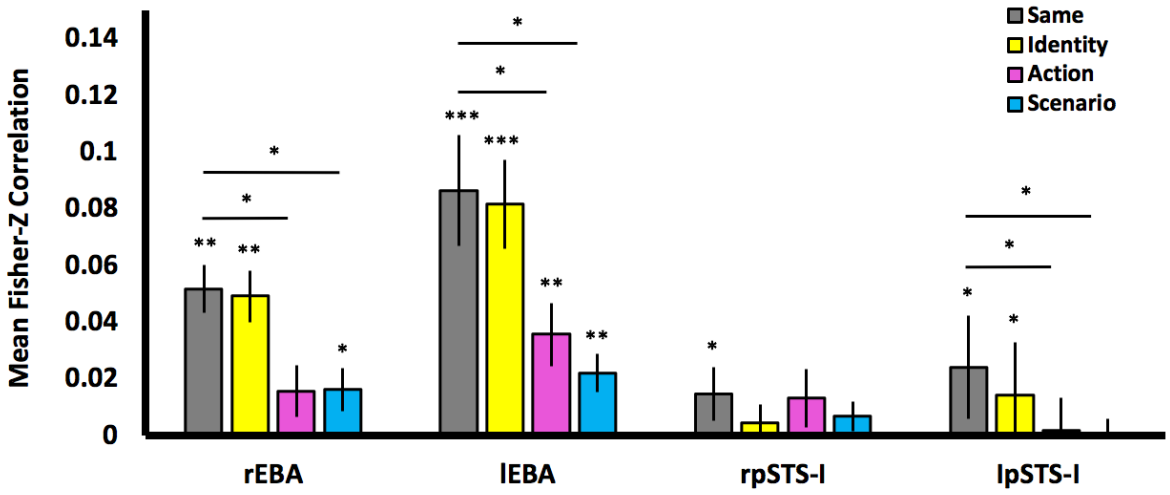


Figure 5.6. A bar chart showing similarity between the dyad RDM and each of the four artificial pairs RDMs. *** = $p \leq .001$; ** = $p \leq .010$; * = $p \leq .05$. Error bars are SEM.

One-sample *t*-tests (uncorrected *ps*) were calculated for the baseline 'same' similarity measure, as only 'above-zero' responses in this measure would allow for interpretable comparisons against the other similarity measures (see figure 5). Both right EBA ($M = .052$, $SD = .083$; $t(17) = 2.66$, $p = .008$) and left EBA ($M = .086$, $SD = .079$; $t(18) = 4.77$, $p < .001$) showed significant responses for the 'same' similarity measure. Similar responses were also shown for the right pSTS-I ($M = .015$, $SD = .037$; $t(19) = 1.76$, $p = .048$) and left pSTS-I ($M = .024$, $SD = .041$; $t(18) = 2.57$, $p = .010$), although these statistical

trends were considerably weaker than for EBA regions. Importantly, both EBA regions showed evidence for strong sensitivity to action information; that is, significant paired *t*-tests for the *same* > *action* (right EBA: $t(17) = 2.51$, $p = .012$; left EBA: $t(18) = 3.75$, $p = .001$) and *same* > *scenario* (right EBA: $t(17) = 2.28$, $p = .018$; left EBA: $t(18) = 4.27$, $p < .001$) contrasts – but crucially, not the *same* > *identity* contrast (right EBA: $t(17) = 0.15$, $p = .441$; left EBA: $t(18) = 0.32$, $p = .377$) were shown. A similar trend was also shown for the left pSTS-I; significant *same* > *action* ($t(18) = 1.81$, $p = .044$) and *same* > *interaction category* ($t(18) = 2.58$, $p = .010$) but not the *same* > *identity* contrast ($t(18) = 1.16$, $p = .132$). However, this trend was not shown for the right pSTS-I – all three contrasts were not significant (*same* > *action*: $t(19) = 0.13$, $p = .450$; *same* > *interaction category*: $t(19) = .085$, $p = .203$; *same* > *identity*: $t(19) = 0.98$, $p = .167$).

These exploratory results suggest that the specific action-gestures performed by interactors contribute significantly to representations of interaction stimuli in the bilateral EBA (and to a lesser extent the left pSTS-I); importantly, these representations remain stable when the identity of one of the interactors is changed, suggesting that these action representations are identity-invariant. Accordingly, these representations do not appear to be strongly affected by the body shape of interactors or idiosyncratic differences in action-gesture execution across interactors.

5.4. Discussion

Overview of Results

The present study aimed to determine whether the pSTS or any other posterior temporal lobe region showed sensitivity to *unique* dyadic information in visually observed interactive scenarios that is not present for isolated individual interactors. Three main findings were shown: 1) Bilateral EBA – but not pSTS – showed evidence consistent with the encoding of specifically dyadic information; 2) The action-gestures of interactors contribute significantly to representations of interactions in the EBA; 3) pSTS classified between three interaction scenarios (i.e. arguing, celebrating, laughing) replicating similar differentiation of types of interactions in moving abstract shapes in (Isik et al., 2017; Walbrin et al., 2018). These findings are discussed below.

Interaction Classification in the pSTS & EBA

Successful classification of the three interactive scenarios (across both dyad and alone stimuli) was observed in the pSTS along with EBA. But which type(s) of information might have contributed to classification? The pSTS is well known to play an important role in biological motion perception (e.g. Deen et al., 2015; Grossman et al., 2000; Pelphrey Morris, Michelich, Allison, & McCarthy, 2005), and is strongly responsive to the presence of contingent movements between interacting figures (Georgescu et al., 2014), as well as dynamic cues that imply interactive behaviour between animate moving shapes (Schultz, Friston, O'Doherty, Wolpert, & Frith, 2005; Gao, Scholl, & McCarthy, 2012). Similarly, the pSTS is also sensitive to the intentional contents that underlie actions (Brass, Schmitt, Spengler, & Gergely, 2007; Pelphrey, Morris, & McCarthy, 2004; Saxe, Xiao, Kovacs, Perrett, & Kanwisher, 2004). It seems plausible that the intentional contents conveyed by dynamic interactive information in both dyad and alone stimuli contributed strongly to pSTS classification.

By contrast the EBA is known to be sensitive to body posture information (e.g. Downing & Peelen, 2011), and is located within the wider LOTC area that is known to be sensitive to diverse representations of action information (Lingnau & Downing, 2015). Although distinct action-gestures were used with each interactive scenario, these tended to be relatively similar to each other (e.g. arguing gestures tended to depict short, sharp movements, while laughing gestures typically contained convulsive movements). Therefore, classification of interaction scenarios in the EBA was likely the result of similar body postures and actions *within* each scenario, that were markedly different *across* the three scenarios.

No Dyadic Information Effect in the pSTS

Despite the pSTS classifying interactive scenarios, the main prediction of this experiment was not supported; no *dyadic information effect* was observed for the pSTS. This contrasts with the findings of Baldassano et al. (2017) that showed an analogous effect in the pSTS (and object selective LOTC) for static depictions of human-object (inter)actions (e.g. a human pushing a cart), compared to averaged responses to isolated objects and humans. One possible explanation concerns STS sensitivity to implied

biological motion in static images (Grossman & Blake, 2001; Peuskens, Vanrie, Verfaillie, & Orban, 2005); static human-object interactions might imply greater biological motion or more effortful movement that is not 'recoverable' from isolated human and objects; for example, an image of a person pushing a cart implies greater movement than the same body pose and cart presented separately, by virtue of greater physical effort required to move the cart, along with the corresponding impression that the cart is moving. Additionally, pSTS sensitivity to causal contingencies (e.g. a billiard ball hitting another, causing a transfer in motion; Blakemore et al., 2001) implicate the potentially strong influence of physical contact in human-object interactions that was not present in the isolated stimuli. By contrast, the current study used dynamic stimuli that contained biological motion information but no physical contact, and as such, the dyad and alone stimuli were closely matched for these two sources of information that might have driven responses to the stimuli used by Baldassano et al. (2017).

Although no dyadic information effect was found in the pSTS, it is important to note that interactive information was still conveyed in the alone stimuli (e.g. communicative gesturing to an unseen interactive partner was strongly implied). Therefore, successful classification of the alone stimuli does not necessarily reflect that pSTS responses are non-interactive. Indeed, in the context of the sorts of gestural interactions used in the current study, it is possible that classification of the alone and dyad stimuli relied on the same cues (i.e. communicative gestures). Similarly, the current data would support the possibility that representations of interactions in this region may encode the presence of two interactors in a linear fashion (i.e. dyad = average of the two individuals). Alternatively, it is possible that the pSTS responses to both dyad and alone stimuli are driven by interactive gestures 'directed' at another individual, regardless of whether the other individual is present or not.

Dyadic Information Processing in the EBA

Although not observed for the pSTS, a dyadic information effect was shown for the EBA, suggesting that the two interactors are encoded in a non-linear manner (i.e. dyads are processed qualitatively differently from the average of two individuals); this effect is considered to be interpretable due to medium, and small-to-medium effect sizes in the right and left hemisphere, respectively (despite not surviving multiple comparison

correction in the right EBA, and only marginal significance at the uncorrected level in the left EBA). Although these effects were not predicted in the EBA, they do fit with previous findings observed in the wider LOTC area. Specifically, Baldassano et al. (2017) observed qualitatively greater responses to human-object interactions than averaged responses to humans and objects in object selective LOTC as well as in the pSTS. However, this trend did not quite reach significance in the EBA, likely due to weaker responses to object stimuli, suggesting that the currently observed EBA responses are specific to dyadic human body information. Similar to the findings of Baldassano et al. (2017), object selective LOTC is sensitive to the spatial configurations of multiple objects that *imply* action (Kaiser & Peelen, 2018; Roberts & Humphreys, 2010).

Broadly, these findings suggest that LOTC regions play a crucial role in the configural processing of *distinct* objects and people from which action representations are formed (e.g. a pitcher placed above a glass implies the action of pouring). In relation to the present findings, it is conceivable that LOTC – and the EBA specifically – performs similar configural processing or grouping of interacting individuals' actions, postures, or movements, into interaction representations. Together, these findings suggest a broad role for the LOTC in constructing (inter)action representations from the spatial relations between multiple objects or people. This proposal is consistent with the broad sensitivity to action information in LOTC (Lingnau & Downing, 2015), and with evidence from the artificial pairs analysis that revealed the important contribution of action information to representations of dyadic interactions in EBA. Therefore, although strong responses were observed in body selective LOTC cortex, these responses undoubtedly contain a strong action information component.

The current study also used dynamic stimuli, in contrast with previous work investigating action grouping responses for static stimuli. Despite evidence that the EBA is highly sensitive to static pose information, and may process body movements as a series of static 'snapshots' (Downing, Peelen, Wiggett, & Tew, 2006; Giese & Poggio, 2003) body (and face) responses are shown to generalise across static and dynamic depictions in broad regions of the posterior temporal cortex (O'Toole et al., 2014). Similarly, representations in the LOTC generalise across dynamic and static depictions of actions (Hafri et al., 2017) and therefore processes underlying action grouping may correspond to relatively abstract representations of action. Indeed, LOTC action representations are invariant to other low-level features, such as movement direction and the specific hand

used to perform an action (Tucciarelli, Turella, Oosterhof, Weisz, & Lingnau, 2015). Therefore, it seems likely that the currently observed dyadic information effect is not inherently dependent on dynamic information.

Instead, consistent with previous findings (Baldassano et al., 2017; Kaiser & Peelen, 2018; Roberts & Humphreys, 2010), it is speculated that dyadic processing of interactions in the LOTC relies on information implied by the *relative interpersonal spatial cues* between the two interactors (e.g. interpersonal distance, physical contact, and facing direction). However, as noted, these previous studies employed static stimuli that did not convey articulated dynamic actions, and so it is possible that grouping of dynamic cues (e.g. temporal correlations between agents, posture changes, and the apparent congruence of) are important to interaction representations in the EBA and LOTC.

Complementary Roles for the EBA and pSTS

It has been suggested that LOTC representations of actions are devoid of 'higher' intentional content (e.g. the motivations for performing a given action) and that full understanding of actions requires further 'intentional elaboration' by other brain regions (Lingnau & Downing, 2015). The pSTS is sensitive to the intentional content underlying observed actions (Brass et al., 2007; Pelphrey et al., 2004; Saxe et al., 2004), and it is possible that such computations are achieved by cross-talk with action processing in LOTC regions such as EBA. Accordingly, it could be speculated that pSTS may aid in the extraction of implicit intentional information (e.g. person A and person B are *arguing*) from relatively decontextualized representations of interactive actions (e.g. person A is leaning forward and shaking a fist while person B is stamping their foot and shaking their head). It is important to note that such intentional processing is likely the product of implicit, spontaneous intentional inference, rather than explicit mental state inference. Indeed, explicit mentalizing drives TPJ responses (e.g. Schurz, Radua, Aichhorn, Richlan, & Perner, 2014), yet the TPJ did not differentiate between interactive scenarios in the present data. It is worth considering that perceived intentionality varies with interactive scenarios, and it is possible that current and previous observations of pSTS modulation to interactive scenarios are at least partly a result of differences in implicit intentional content.

Strengths & Limitations

The current findings are supported by several notable methodological strengths. By using a cross-classification approach, direct comparison of whether response patterns to dyad and alone stimuli reliably differed was possible; this is contrasted with univariate comparisons that cannot reliably determine whether responses are the same, even if mean responses do not differ (i.e. qualitatively different patterns of activation could yield similar mean responses). Therefore, this analysis allowed for the detection of subtle dyadic information effects. Additionally, the novel implementation of the artificial pairs analysis allowed for the systematic testing of how different types of social information (e.g. identity) contribute to dyadic representations.

Along with these strengths, there are also several notable limitations with the current research. As mentioned previously, the three interactive scenarios were specifically chosen to be as distinct as possible to allow for relatively strong classification, to test for dyadic information effects. However, it is difficult to know exactly which types of information contributed to classification in the EBA and pSTS and it is likely that differences in emotion and valence information partially contributed to classification of these categories; however, given that different emotions featured within each scenario (e.g. anger and frustration for arguing), it is unlikely that classification was predominantly driven by emotion specifically. However, large differences in valence information (e.g. between arguing and celebrating) may have contributed to classification performance. Additionally, both the classification and artificial pairs analysis revealed relatively subtle effects in the EBA (i.e. searchlight analysis did not show any surviving clusters for the dyadic information effect, and artificial pairs analysis revealed small correlation values). However, these nuanced effects were highly similar in both hemispheres, and were substantially different to responses in other areas, suggesting that they are reliable and specific to EBA.

Conclusions

In summary, the present results show that representations of dyadic social interactions in the EBA are sensitive to dyadic information beyond that which is encoded by the simple average of two separate interactors presented in isolation. This so-called

dyadic information effect, suggests that the EBA plays an important role in forming an interactive context from the actions and spatial relations between two individuals. These findings also show that both EBA and pSTS differentiate between different types of social interactions, and suggests potentially complementary roles in human social interaction perception for these two proximal regions.

Chapter 6

General Discussion

6.1. Overview of Findings

Several insights about visual brain responses to social interactions are observed across the experiments presented in this thesis. In chapter 3, it was demonstrated that the right posterior superior temporal sulcus (pSTS) was virtually the only region to discriminate interacting human point-light figures from non-interacting figures, suggesting a crucial role for the pSTS in detecting the presence of social interactions. Responses in this functionally localized region were then tested in a separate group of subjects who viewed non-human moving shape stimuli. Using a support vector machine (SVM) classification approach, the right pSTS significantly differentiated interactions from non-interactions. The pSTS also demonstrated sensitivity to the type of interactive scenario – that is, whether the moving shapes were competing or cooperating with each other. Similar, although weaker, responses were also observed in neighbouring right temporoparietal junction (TPJ) cortex, suggesting that the pSTS, and to a lesser extent, the TPJ, detect and differentiate between abstract depictions of social interactions. A third 'control' region – within the lateral occipital temporal cortex (LOTC) – also differentiated the interaction and non-interaction conditions, but not the competition and cooperation conditions, suggesting action specific, rather than interaction specific processing in this region. Additionally, stronger overall univariate responses in the pSTS than TPJ further implicate the central role that the pSTS plays in visual social interaction perception.

In chapter 4, responses to interacting point-light figures were tested in both adults and pre-adolescent children within functionally localized pSTS, as well as in other 'social' temporal lobe regions. This allowed for the comparison of interaction responses within multiple functionally localized regions, as well as responses to other types of social information in the pSTS; this in turn allowed for the characterization of differences in social interaction brain responses between children and adults. Crucially, support for the main hypothesis was shown: Adults showed greater interaction selectivity than children in the pSTS. This difference was large in the left pSTS, but only marginal in the right pSTS, where children's responses were weaker but comparable to adults. Unlike children, adults showed significant interaction selectivity in neighbouring regions of the STS too. Further analyses showed that adults' interaction selective responses in the pSTS are more 'focally-tuned' than in children, who showed weaker, more diffuse selectivity. These results reveal that selective responses to visually observed social interactions are not

fully mature in pre-adolescent children, and undergo further development across adolescence.

In chapter 5, live-action human interaction video clips were used to investigate whether the pSTS or other 'social' posterior temporal lobe regions differentiate between different types of expressive human interactions, and crucially, whether these regions demonstrate a *dyadic information effect*; that is, that neural responses to dyadic social interactions contain *unique* information that cannot be 'recovered' by the averaged responses to the constituent interactors, presented in isolation. It was found that the bilateral pSTS and EBA were the only regions that reliably differentiated between interactions. However, against the main hypothesis, the pSTS did not show evidence consistent with a dyadic information effect; instead, this effect was observed in right EBA (and to a lesser extent, left EBA). Follow-up analyses showed that interaction representations in the EBA are most sensitive to disruptions in the specific action-gestures performed by interactors; this suggests that the action information conveyed by each interactor contributes strongly to interaction representations in the EBA. Together, these findings suggest that both the pSTS and EBA play important roles in the visual perception of dyadic human interactions; EBA responses implicate this region (and potentially wider LOTC cortex) in the dyadic processing of individuals during interaction perception, and it is speculated that the pSTS plays a complementary role in extracting the intentional contents of interactive scenarios.

6.2. Synthesis of Findings Across Experiments

Across the three experiments presented here, the pSTS demonstrates sensitivity to both the presence of an interaction (compared to two non-interacting individuals) and sensitivity to different types of interactions. These findings provide a compelling demonstration of the importance of the pSTS in processing visually observed dyadic interactions. The EBA was shown to reliably differentiate between human body interactions and appears to play an important role in the integrative processing of dyadic action behaviour (similarly, a neighbouring region of LOTC showed a response pattern consistent with sensitivity to action information in moving shape interactions). However, despite strong univariate responses to body information, the EBA does not appear to make univariate distinctions between interactions and non-interactions in the same way

that the pSTS does. These findings suggest that the two regions make differential – and likely complementary – contributions to visual social interaction processing.

Although the TPJ showed similar, albeit it weaker, responses to social interactions portrayed by abstract moving shapes, such responses were absent when viewing other interaction stimuli. This suggests that responses to the abstract moving shape stimuli may reflect additional processing required to recognize and differentiate interactions in the absence of human body information. Similarly, despite considerable interaction selectivity in the face selective region of the STS (STS-F) in chapter 4, this region did not differentiate between different human interactions (in which high-spatial frequency face information was removed) in chapter 5. Therefore, the findings of this thesis suggest that TPJ and face selective STS may contribute to social interaction perception in certain viewing conditions only.

6.3. Why is the pSTS Important for Processing Third Person Interactions?

There are two characteristics of the pSTS that suggest it might be 'optimized' for processing third-person social interactions. Firstly, the pSTS and neighbouring cortex is known to process numerous types of social information that typically contribute to interaction perception such as face, body, biological motion, mentalizing, and voice information (Lahnakoski et al., 2012; Deen, Koldewyn, Kanwisher, & Saxe, 2015). The functional organization of the STS is complex, and is comprised of overlapping regions that are selective for different socially relevant categories of information (Deen et al., 2015). Indeed, functional connectivity data suggests that the pSTS may constitute distinct regions that are differentially connected to other brain regions to serve different visual social processes (Shih et al., 2011). These accounts are not exclusive of each other, and it seems likely that diverse social processing across distinct but neighbouring STS regions contributes to social interaction processing in the pSTS.

Secondly, responses in the wider posterior temporal cortex are more sensitive to third-person than first-person social information. Visual responses to allocentric > egocentric depictions of hand actions are shown in the right temporo-parietal cortex (Wurm, von Cramon, & Schubotz, 2011; Wurm & Schubotz, 2018), and for static body parts in the right EBA (Chan, Peelen, & Downing, 2004; Saxe, Jamal, & Powell, 2005). Similarly, greater posterior middle and superior temporal cortex responses are shown

when observing simple actions compared to when performing them (Committeri et al., 2015; Noordzij et al., 2010). This broad distinction may even extend to linguistic processing as greater responses in left pSTS and hMT+ are observed when contrasting third-person > first-person verbs (Papeo & Lingnau, 2015). Given the strong sensitivity to third-person information and rich social processing across the STS, the location of the pSTS may be 'ideally suited' to processing third-person social interactions.

6.4. Complementary Functions of pSTS & EBA/LOTC

Several accounts of the complementary roles of pSTS and EBA/LOTC are proposed below. These accounts are largely speculative, but aim to consolidate the current findings as well as motivate further research to understand how these regions, together, contribute to social interaction perception.

6.4.1. Body Motion & Body Form Processing

The prominent role that the pSTS plays in observing dynamic social interactions is unsurprising, given robust responses to biological motion in and around this area (Allison, Puce, & McCarthy, 2000; Grossman et al., 2000; Pelphrey Morris, Michelich, Allison, & McCarthy, 2005). However, greater responses to interaction than non-interaction stimuli that also contained two individuals' articulated movements (chapters 3 & 4), demonstrate that this region is not simply responsive to biological motion, but also the perception of interactive information that arises from these stimuli. Previous studies have shown pSTS modulation when manipulating motion parameters such as the contingency of movements within an interacting dyad (Georgescu et al., 2014) and the degree of correlated movement between interacting shapes (Schultz, Friston, O'Doherty, Wolpert, & Frith, 2005). Movement contingency may represent one of the most powerful dynamic social interaction cues, as the movements of a given interactor are strongly influenced by – and exert influence on – the movements of other interactors, demonstrating the fundamentally dynamic, reactive nature of interactions (De Jaeger, Di Paolo, & Gallagher, 2010). Indeed, strong pSTS responses were observed when contrasting interactions comprised of contingent actions relative to similar scenarios

with non-contingent actions (chapters 3 & 4), and this cue may potentially represent the dominant motion characteristic driving pSTS responses in the present data.

In chapter 5, evidence consistent with dyadic processing of dynamic interactors is shown in the EBA. However, it has been proposed that body representations are encoded as a series of static 'snapshots', and that this region is not preferentially sensitive to dynamic information per se. Accordingly, Giese and Poggio (2003) propose a model of biological motion recognition that consists of two distinct (yet interacting) hierarchically organised processing streams for body motion and body form. Each stream emerges from simple orientation or local motion detectors in early visual areas and terminates with complex biological motion processing in the STS. Specifically, 'snapshot' neurons that encode body shape in inferior temporal regions such as EBA, and motion pattern detectors in neighbouring motion sensitive cortex (e.g. motion sensitive middle temporal cortex (MT)) pass information from these respective form and motion streams, onto higher level motion sensitive neurons in the STS that represent a convergence point between the two pathways.

This proposed distinction is supported by Downing, Peelen, Wiggett, and Tew (2006); greater EBA sensitivity was observed for briefly presented sequences of highly variable 'incongruent' poses of single bodies, compared to sequences of highly similar 'congruent' poses, whereas the opposite trend was observed in the pSTS. It was concluded that EBA represents body posture in a 'snapshot' fashion, unlike the pSTS that favours smooth, continuous body motion. Complementary causal dissociations have also been inferred using transcranial magnetic stimulation (TMS), in which stimulation over EBA selectively disrupts static body perception, whereas pSTS stimulation perturbs dynamic body perception (Vangeneugden, Peelen, Tadin, & Battelli, 2014). Analogous results have also been shown for static and dynamic face perception when stimulating face selective cortex that neighbours EBA (i.e. occipital face area) and pSTS, respectively (Pitcher, Duchaine, & Walsh, 2014). Based on these findings, it seems plausible that dynamic interaction sensitivity in the pSTS arises from either motion cues alone, or with the contribution of body form based 'snapshot' representations within the EBA.

6.4.2. Differential Action Understanding

In contrast to a body specific account of interaction perception, it was demonstrated in chapter 3 that interactive behaviour can be recognized and understood via abstract moving shapes that do not contain human body information. Additionally, a functional distinction between the pSTS and a region of LOTC – in close proximity to (and likely overlapping with) EBA – was observed. Specifically, pSTS is sensitive to both the presence and contents of abstract moving shape interactions (i.e. above-chance classification for both the interaction vs. non-interaction, and competition vs. cooperation contrasts), whereas LOTC is apparently only sensitive to large differences in the action contents of these stimuli.

The broader LOTC area is involved in processing multiple aspects of action information, such as body perception, motion perception, tool perception, action performance, along with semantic and conceptual action information (Lingnau & Downing, 2015). It has also been proposed that different dimensions of action information are organized in a graded fashion across this region, for example, social and object-directed actions are more strongly represented in dorsal and ventral areas respectively, while concrete-abstract representations show a posterior-anterior organization (Wurm, Caramazza, & Lingnau, 2017).

Another important aspect of visual LOTC responses is their apparent invariance across low-level features. For example, observed actions are shown to generalise across static and moving depictions of the same actions (Hafri, Trueswell, & Epstein, 2017), and across movement direction and the specific hand used to perform an action (Tucciarelli, Turella, Oosterhof, Weisz, & Lingnau, 2015). As such, action representations in this region are relatively *abstract* and may serve a general aim to '… change the state of the world in some way' rather than conveying complex intentional information that is likely subserved by higher level regions (Lingnau & Downing, 2015).

As such, action representations within the LOTC may constitute an 'intermediate level' of action understanding. For example, an observed interaction might be represented as a coherent action in the LOTC ('two people are shaking hands'), rather than a more basic ('person A is extending their right hand to shake the right hand of person B') or more contextualised higher level understanding ('the two people are greeting each other'). It is also likely that, given the key role of the pSTS in understanding

the immediate intentions that underlie observed actions (Brass, Schmitt, Spengler, & Gergely, 2007; Pelphrey, Morris, & McCarthy, 2004; Saxe, Xiao, Kovacs, Perrett, & Kanwisher, 2004; Shultz, Lee, Pelphrey, & McCarthy, 2010), this region may process the higher level intentional contents of actions.

6.4.3. Combined Account

A final account is now proposed that combines elements of the two preceding suggestions. The pSTS is implicated as a social 'hub' that coactivates with task specific brain networks that vary across different visual social stimuli (Dasgupta, Tyler, Wicks, Srinivasan, & Grossman, 2017; Lahnakoski et al., 2012), and may constitute a convergence point between person perception, action observation, and mentalizing networks (Yang, Rosenblau, Keifer, & Pelphrey, 2015; Quadflieg & Koldewyn, 2017). One possible corollary of this evidence is that social interaction detection is 'constructed' in the pSTS from numerous streams of information across different brain regions, with crucial contributions from LOTC regions. In addition to pSTS and LOTC occupying neighbouring regions of the wider posterior temporal cortex, strong functional coupling between these regions is shown when viewing various visual social stimuli (Dasgupta et al., 2017; Nath & Beauchamp, 2012; Shih et al., 2011; Van den Stock, Hortensius, Sinke, Goebel, & de Gelder, 2015).

The functional coupling of these two regions is further underscored by the complementarity of processing tendencies; sub-regions of the LOTC are shown to be sensitive to spatial configurations of static object and human arrangements that show or imply action behaviour (Baldassano, Beck, & Fei-Fei, 2017; Kaiser & Peelen, 2018; Roberts & Humphreys, 2010) and are reasoned to process information in a 'snapshot' manner (Downing et al., 2006; Giese & Poggio, 2003). By contrast, the pSTS is also known to be sensitive to continuous sequences of coherent social movement (Downing et al., 2006; Pitcher, Dilks, Saxe, Triantafyllou, & Kanwisher, 2011) and intentional understanding from dynamic actions (Brass et al., 2007; Pelphrey et al., 2004; Saxe et al., 2004; Shultz et al., 2010).

Finally, one possible explanation of how these two regions share information is as follows. Transient action 'snapshots' in LOTC regions are generated by integrating body postural and spatial-relational information conveyed by interactors; in the pSTS,

intentional or contextual understanding arises and is modulated by the cumulative stream of action snapshots, as well as by direct detection in the motion processing stream. This account describes social interaction viewing under normal viewing conditions, however, it remains to be seen to whether dynamic social interaction perception is preserved when LOTC functioning is disrupted, as is shown for dynamic body and face processing (Pitcher et al., 2014; Vangeneugden et al., 2014). Although speculative, this account gives a plausible explanation for how the pSTS and LOTC regions might together contribute to social interaction perception.

6.5. Core Mentalizing and Action Observation Network Contributions to Social Interaction Perception

Along with pSTS and LOTC, other brain regions undoubtedly contribute to naturalistic social interaction processing. In terms of the mentalizing network, no medial prefrontal cortex (mPFC) responses were found for any of the current experiments, likely due to the use of orthogonal response tasks; by contrast mPFC activation is typically found in social interaction studies that require explicit social judgements (Centelles, Assaiante, Nazarian, Anton, & Schmitz, 2011; Dolcos, Sung, Argo, Flor-Henry, & Dolcos, 2012; Wang & Quadflieg, 2015). TPJ demonstrated some sensitivity to moving shape interactions, but not point-light or live-action human stimuli; this suggests that the TPJ does not play a general role in interaction processing, but may be recruited in abstract scenarios where human body or face information is absent. Additionally, responses in the precuneus were, observed for adults in chapter 4, but surprisingly, not for the same contrast in chapter 3. As discussed in chapter 4, this region may play a role in perceiving the congruence of interactors movements (Petrini, Piwek, Crabbe, Pollick, & Garrod, 2014), or directional information (Carlin, Calder, Kriegeskorte, Nili, & Rowe, 2011) – and plausibly facing direction – depicted by interactors. However, these possibilities seem unlikely given the absence of differentiation in the precuneus for congruency and facing direction comparisons with dyadic static stimuli (Quadflieg, Gentile, & Rossion, 2015). Instead, precuneus activation may reflect more general spatial attentional processing (e.g. Culham, Brandt, Cavanagh, Kanwisher, Dale, & Tootell, 1998; Le, Pardo, & Hu, 1998).

Core action observation network (AON) regions did not contribute strongly in any of the present experiments; no inferior parietal lobule (IPL) responses were shown,

although a very small inferior frontal gyrus (IFG) cluster was found in the whole brain analysis in chapter 4, for adults only (but not for the same contrast in chapter 3). That these regions did not contribute strongly is perhaps unsurprising, given the traditional interpretation of responses in these regions (i.e. 'mirroring' or simulating others' actions; e.g. Van Overwalle & Baertens, 2009). Additionally, this is supported by the proposal that superior temporal gyrus (STG) regions play a more prominent role in action understanding when the observed actions cannot be easily simulated by a viewer (Van Overwalle & Baertens, 2009), as is likely the case when viewing multiple, simultaneous interactors. IFG regions are also shown to be more sensitive to first-person than third-person action information (Oosterhof, Tipper, & Downing, 2012; Papeo, Corradi-Dell'Acqua, & Rumiati, 2011), conforming to the notion that IFG is involved in motor simulation of the actions performed by an individual. However, previous interaction studies often report IFG activation, and various explanations such as the presence of goal-oriented behaviour (Canessa et al., 2012), general motor simulation (Centelles et al., 2011), and action-intention understanding (Georgescu et al., 2014) are proposed.

IPL (or superior parietal lobule) responses have been reported in other social interaction studies (e.g. Dolcos et al. 2012; Kujala et al., 2012a). Interestingly, it has been speculated that parietal responses may correspond to processing related to the interpersonal distance between interactors (Quadflieg & Koldewyn, 2017) – a suggestion that complements evidence for egocentric distance computations in the IPL (Parkinson, Liu, & Wheatley, 2014), and might suggests close functional coupling with spatial configural processing in the LOTC. Although clear contributions of mentalizing and AON regions were not evident in the present data, subsequent research may serve to better characterize exactly which information these regions are sensitive to during social interaction observation.

6.6. Overlapping Constructs to Social Interaction

Several constructs that are non-orthogonal to social interactions and are known to modulate pSTS responses are worth noting. pSTS (and wider STS) is sensitive to a variety of communicative signals (Redcay, 2008), depicted by written text (Redcay, Velnoskey, & Rowe, 2016), hand gestures (Noordzij et al., 2010; Saggar, Shelly, Lepage, Hoeft, & Reiss, 2014; Yang, Andric, & Matthew, 2015), and speech information (Röder,

Stock, Neville, Bien, & Rosler, 2002; Shultz, Vouloumanos, & Pelphrey, 2012; Wise et al., 2001). Similarly, the pSTS is strongly sensitive to the immediate intentions of observed actions (Brass et al., 2007; Pelphrey et al., 2004; Saxe et al., 2004; Shultz et al., 2010), and strong, modulated responses are observed when explicit, deliberative inferences are attributed to randomly moving shapes (Lee, Gao, & McCarthy, 2012). Although communicative signalling and understanding action-intentions undoubtedly contribute to interaction processing in many scenarios, responses to moving shape interactions in chapter 3 suggest that interaction sensitivity cannot be fully explained by these factors. For example, these scenarios did not convey explicit communicative information, and although broad differences in intentionality ratings might have confounded the interaction vs. non-interaction contrast, this was not the case with competition vs. cooperation contrast (i.e. intentionality ratings were matched between these conditions). These non-orthogonal sources of social information are often present during social interactions, and as such, likely contribute to understanding social interactions. This further supports the notion that the pSTS is a hub that receives rich visual, auditory, and semantic information, necessary for the immediate recognition and categorization of social interactions.

6.7. Methodological Strengths and Limitations

The novel contributions of the present findings are emphasized by several methodological advantages that have often been absent from previous social interaction research. Firstly, a functional localization approach was employed, and as such, regions were defined on the basis of their 'preferred' response categories rather than by arbitrary anatomical definitions. This is important, given the wide inter-subject variability in location of functionally specific regions (Saxe, Brett, & Kanwisher, 2006), especially when considering the complex spatial organization of functional responses in areas like the STS (Deen et al., 2015). In addition to the high functional specificity achieved with ROI analyses, 'contamination' of top-down mentalizing was minimized by using orthogonal response tasks, and visual responses to social interactions were better isolated from these top-down processes. Additionally, the use of multivariate approaches (i.e. SVM classification and representational similarity analysis) allowed for sensitive discriminations between conditions that are not always possible with univariate analyses

(e.g. competition and cooperation are qualitatively different interactions that might not be quantitatively differentiated by univariate analysis).

Although similar stimuli were used to localize interaction selective pSTS, the use of 3 different stimulus sets to test interaction responses show that this region is sensitive to visually diverse depictions of interactive behaviour. Moreover, this general interaction sensitivity in the pSTS is underscored by relatively more restricted responses in other social regions. For example, the TPJ was only sensitive to interactive scenarios depicted by moving shape stimuli.

It is also worth considering a few limitations of the current research. The use of an ROI-led approach across experiments allowed for the detection of subtle, specific effects, that were partly supported with less sensitive whole brain analyses. However, weaker responses in other brain regions that are not detectable with whole brain analysis, and hence omitted from ROI analyses that might show detectable effects, have potentially been overlooked.

Despite converging evidence for social interaction sensitivity in pSTS, specific interaction cues were not systematically disrupted. For example, interactions were contrasted with different control stimuli that varied by several cues such as facing direction, the presence of object-directed actions, and the presence of an interaction partner. Subsequent research should aim to isolate and test the contributions of these signals to determine which cues are most informative, and whether they are linearly weighted across different viewing scenarios.

6.8. Further Research

The presently undertaken research provides a basis for visual third-person interaction perception in the posterior temporal cortex. The following outstanding research questions may serve to guide future research aimed at fully characterizing social interaction responses in the brain.

- What are the dominant temporal and spatial interaction cues, and where are they encoded in the brain?
    - Does perturbing the temporal contingencies between interacting individuals' actions reduce pSTS responses?

    - How does the relative proximity of interactors change neural responses to interactions?
- Are pSTS interaction responses multimodal (e.g. similar across visual and auditory scenarios)?
    - Do neural responses generalise or cross-classify across modalities?
- How does the pSTS respond to more complex interactions containing more than two individuals?
    - Do triadic interactions evoke stronger neural responses than dyadic interactions?
- Which network of regions functionally coactivate with the pSTS during social interactions?
    - Is this modulated by task (e.g. do explicit, inferential tasks drive stronger correlations between temporal lobe regions and mPFC)?
- Which regions encode viewpoint invariant depictions of interactions?
- How does interpersonal distance between interactors contribute to interaction perception?

6.9. Conclusion

Converging findings across three fMRI experiments show the importance of the posterior temporal cortex in visually perceiving third-person social interactions. The pSTS is strongly implicated as a key region in recognising the presence of interactions and differentiating between qualitatively different interactive scenarios. Processing of action contents and the integration of interactors' actions in neighbouring EBA (and potentially other LOTC regions) also make important contributions to these perceptual interaction processes. As with other social categories of information, such as faces, bodies, and theory of mind, neural responses to social interactions are not fully mature in pre-adolescent children, suggesting further neural development across adolescence. The importance of these findings rests upon the demonstration that social brain regions that have traditionally been targeted with individual human stimuli (e.g. individual human bodies in pSTS and EBA) are sensitive to the contents of social interactive behaviour conveyed by multiple individuals. Together, these findings provide a strong basis to further examine posterior temporal responses to social interactions, and future

studies focused on social interaction perception may also serve to inform social information processing deficits in clinical populations, such as autism and schizophrenia.

References

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences*, *4*(7), 267-278.

Ardekani, B. A., Bachman, A. H., & Helpern, J. A. (2001). A quantitative comparison of motion detection algorithms in fMRI. *Magnetic Resonance Imaging*, *19*(7), 959-963.

Avidan, G., & Behrmann, M. (2009). Functional MRI reveals compromised neural integrity of the face processing network in congenital prosopagnosia. *Current Biology*, *19*(13), 1146-1150.

Balas, B., Kanwisher, N., & Saxe, R. (2012). Thin-slice perception develops slowly. *Journal of Experimental Child Psychology*, *112*(2), 257-264.

Baldassano, C., Beck, D. M., & Fei-Fei, L. (2017). Human-object interactions are more than the sum of their parts. *Cerebral Cortex*, *27*(3), 2276-2288.

Bandettini, P. A., & Cox, R. W. (2000). Event-related fMRI contrast when using constant interstimulus interval: Theory and experiment. *Magnetic Resonance in Medicine*, *43*(4), 540-548.

Bandettini, P. A., Wong, E. C., Hinks, R. S., Tikofsky, R. S., & Hyde, J. S. (1992). Time course EPI of human brain function during task activation. *Magnetic Resonance in Medicine*, *25*(2), 390-397.

Barraclough, N. E., Xiao, D., Baker, C. I., Oram, M. W., & Perrett, D. I. (2005). Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *Journal of Cognitive Neuroscience*, *17*(3), 377-391.

Bassili, J. N. (1976). Temporal and spatial contingencies in the perception of social events. *Journal of Personality and Social Psychology*, *33*(6), 680.

Batty, M., & Taylor, M. J. (2006). The development of emotional face processing during childhood. *Developmental Science*, *9*(2), 207-220.

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, *41*(5), 809-823.

Beauchamp, M. S., Nath, A. R., & Pasalar, S. (2010). fMRI-guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *Journal of Neuroscience*, *30*(7), 2414-2417.

Belin, P., Zatorre, R. J., Lafallie, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 309.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*, 289-300.

Bertenthal, B. I., Proffitt, D. R., & Kramer, S. J. (1987). Perception of biomechanical motions by infants: Implementation of various processing constraints. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(4), 577.

Blakemore, S. J., Fonlupt, P., Pachot-Clouard, M., Darmon, C., Boyer, P., Meltzoff, A. N., … Decety, J. (2001). How the brain perceives causality: An event-related fMRI study. *NeuroReport*, *12*(17), 3741-3746.

Bonte, M., Frost, M. A., Rutten, S., Ley, A., Formisano, E., & Goebel, R. (2013). Development from childhood to adulthood increases morphological and functional inter-individual variability in the right superior temporal cortex. *NeuroImage*, *83*, 739-750.

Brass, M., Schmitt, R. M., Spengler, S., & Gergely, G. (2007). Investigating action understanding: Inferential processes versus action simulation. *Current Biology*, *17*(24), 2117-2121.

Brett, M., Anton, J. L., Valabregue, R., & Poline, J. B. (2002, June). *Region of interest analysis using an SPM toolbox.* Paper presented at 8th International Conference on Functional Mapping of the Human Brain, Sendai, Japan. Retrieved from: https://matthew.dynevor.org/research/abstracts/marsbar/marsbar_abstract.pdf

Brey, E., & Shutts, K. (2015). Children use nonverbal cues to make inferences about social power. *Child Development*, *86*, 276-286.

Burock, M. A., Buckner, R. L., Woldorff, M. G., Rosen, B. R., & Dale, A. M. (1998). Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. *NeuroReport*, *9*(16), 3735-3739.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, *10*(11), 649-657.

Canessa, N., Alemanno, F., Riva, F., Zani, A., Proverbio, A. M., Mannara, N., … Cappa, S. F. (2012). The neural bases of social intention understanding: The role of interaction goals. *PLoS ONE*, *7*(7), e42347.

Carlin, J. D., Calder, A. J., Kriegeskorte, N., Nili, H., & Rowe, J. B. (2011). A head view-invariant representation of gaze direction in anterior superior temporal sulcus. *Current Biology*, *21*(21), 1817-1821.

Carter, E. J., & Pelphrey, K. A. (2006). School-aged children exhibit domain-specific responses to biological motion. *Social Neuroscience*, *1*(3), 396-411.

Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *NeuroImage*, *50*(3), 1148-1167.

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

Centelles, L., Assaiante, C., Etchegoyhen, K., Bouvard, M., & Schmitz, C. (2013). From action to interaction: Exploring the contribution of body motion cues to social understanding in typical development and in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *43*(5), 1140-1150.

Centelles, L., Assaiante, C., Nazarian, B., Anton, J. L., & Schmitz, C. (2011). Recruitment of both the mirror and the mentalizing networks when observing social interactions depicted by point-lights: A neuroimaging study. *PLOS One*, *6*, e15749.

Chan, A. W., Peelen, M. V., & Downing, P. E. (2004). The effect of viewpoint on body representation in the extrastriate body area. *NeuroReport*, *15*(15), 2407-2410.

Chang, C. C., & Lin, C. J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *2*(3), 27.

Committeri, G., Cirillo, S., Costantini, M., Galati, G., Romani, G. L., & Aureli, T. (2015). Brain activity modulation during the production of imperative and declarative pointing. *NeuroImage*, *109*, 449-457.

Coutanche, M. N., Solomon, S. H., & Thompson-Schill, S. L. (2016). A meta-analysis of fMRI decoding: Quantifying influences on human visual population codes. *Neuropsychologia*, *82*, 134-141.

Cross, E. S., Hamilton, A. F. D. C., Kraemer, D. J., Kelley, W. M., & Grafton, S. T. (2009). Dissociable substrates for body motion and physical experience in the human action observation network. *European Journal of Neuroscience*, *30*(7), 1383-1392.

Culham, J. C., Brandt, S. A., Cavanagh, P., Kanwisher, N. G., Dale, A. M., & Tootell, R. B. (1998). Cortical fMRI activation produced by attentive tracking of moving targets. *Journal of Neurophysiology*, *80*(5), 2657-2670.

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*(2), 158-164.

Dasgupta, S., Tyler, S. C., Wicks, J., Srinivasan, R., & Grossman, E. D. (2017). Network connectivity of the right STS in three social perception localizers. *Journal of Cognitive Neuroscience*, *29*(2), 221-234.

Op de Beeck, H. P. O., Haushofer, J., & Kanwisher, N. G. (2008). Interpreting fMRI data: Maps, modules and dimensions. *Nature Reviews Neuroscience*, *9*(2), 123-135.

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition?. *Trends in Cognitive Sciences*, *14*(10), 441-447.

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

Deen, B., Koldewyn, K., Kanwisher, N., & Saxe, R. (2015). Functional organization of social perception and cognition in the superior temporal sulcus. *Cerebral Cortex*, *25*(11), 4596-4609.

Deen, B., Richardson, H., Dilks, D. D., Takahashi, A., Keil, B., Wald, L. L., ... & Saxe, R. (2017). Organization of high-level visual cortex in human infants. *Nature Communications*, *8*, 13995.

Deen, B., & Saxe, R. (2012). Neural correlates of social perception: The posterior superior temporal sulcus is modulated by action rationality, but not animacy. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *34*(34), 276-281.

Ding, X., Gao, Z., & Shen, M. (2017). Two equals one: Two human actions during social interaction are grouped as one unit in working memory. *Psychological Science*, *28*(9), 1311-1320.

Dodell-Feder, D., Koster-Hale, J., Bedny, M., & Saxe, R. (2011). fMRI item analysis in a theory of mind task. *NeuroImage*, *55*(2), 705-712.

Dolcos, S., Sung, K., Argo, J. J., Flor-Henry, S., & Dolcos, F. (2012). The power of a handshake: Neural correlates of evaluative judgments in observed social interactions. *Journal of Cognitive Neuroscience*, *24*(12), 2292-2305.

Downing, P. E., Chan, A. Y., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2005). Domain specificity in visual cortex. *Cerebral Cortex*, *16*(10), 1453-1461.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*(5539), 2470-2473.

Downing, P. E., & Peelen, M. V. (2011). How might occipitotemporal body selective regions interact with other brain areas to support person perception? *Cognitive Neuroscience*, *2*(3), 216-226.

Downing, P. E., Peelen, M. V., Wiggett, A. J., & Tew, B. D. (2006). The role of the extrastriate body area in action perception. *Social Neuroscience*, *1*, 52-62.

Duchaine, B., & Yovel, G. (2015). A revised neural framework for face processing. *Annual Review of Vision Science*, *1*, 393-416.

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, *113*(28), 7900-7905.

Fabbri-Destro, M., & Rizzolatti, G. (2008). Mirror neurons and mirror systems in monkeys and humans. *Physiology*, *23*(3), 171-179.

Fawcett, C., & Gredebäck, G. (2013). Infants use social context to bind actions into a collaborative sequence. *Developmental Science*, *16*(6), 841-849.

Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magnetic Resonance in Medicine*, *33*(5), 636-647.

Freire, A., Lewis, T. L., Maurer, D., & Blake, R. (2006). The development of sensitivity to biological motion in noise. *Perception*, *35*(5), 647-657.

Gallagher, S. (2004). Understanding interpersonal problems in autism: Interaction theory as an alternative to theory of mind. *Philosophy, Psychiatry, & Psychology*, *11*(3), 199-217.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*(2), 593-609.

Gao, T., McCarthy, G., & Scholl, B. J. (2010). The wolfpack effect: Perception of animacy irresistibly influences interactive behavior. *Psychological Science*, *21*(12), 1845-1853.

Gao, T., Newman, G. E., & Scholl, B. J. (2009). The psychophysics of chasing: A case study in the perception of animacy. *Cognitive Psychology*, *59*(2), 154-179.

Gao, T., Scholl, B. J., & McCarthy, G. (2012). Dissociating the detection of intentionality from animacy in the right posterior superior temporal sulcus. *Journal of Neuroscience*, *32*(41), 14276-14280.

Georgescu, A. L., Kuzmanovic, B., Santos, N. S., Tepest, R., Bente, G., Tittgemeyer, M., & Vogeley, K. (2014). Perceiving nonverbal behavior: Neural correlates of processing movement fluency and contingency in dyadic interactions. *Human Brain Mapping*, *35*(4), 1362-1378.

Giese, M. A., & Poggio, T. (2003). Cognitive neuroscience: Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience*, *4*(3), 179.

Gläscher, J. (2009). Visualization of group inference data in functional neuroimaging. *Neuroinformatics*, *7*(1), 73-82.

Gobbini, M. I., Koralek, A. C., Bryan, R. E., Montgomery, K. J., & Haxby, J. V. (2007). Two takes on the social brain: A comparison of theory of mind tasks. *Journal of Cognitive Neuroscience*, *19*(11), 1803-1814.

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., ... & Rapoport, J. L. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences*, *101*(21), 8174-8179.

Golarai, G., Ghahremani, D. G., Whitfield-Gabrieli, S., Reiss, A., Eberhardt, J. L., Gabrieli, J. D., & Grill-Spector, K. (2007). Differential development of high-level visual cortex correlates with category-specific recognition memory. *Nature Neuroscience*, *10*(4), 512.

Gray, K. L., Barber, L., Murphy, J., & Cook, R. (2017). Social interaction contexts bias the perceived expressions of interactants. *Emotion*, *17*(4), 567.

Grezes, J., Pichon, S., & De Gelder, B. (2007). Perceiving fear in dynamic body expressions. *NeuroImage*, *35*(2), 959-967.

Grossman, E. D., & Blake, R. (2001). Brain activity evoked by inverted and imagined biological motion. *Vision Research*, *41*(10), 1475-1482.

Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., & Blake, R. (2000). Brain areas involved in perception of biological motion. *Journal of Cognitive Neuroscience*, *12*(5), 711-720.

Hadad, B. S., Maurer, D., & Lewis, T. L. (2011). Long trajectory for the development of sensitivity to global and biological motion. *Developmental Science*, *14*(6), 1330-1339.

Hadjikhani, N., & De Gelder, B. (2002). Neural basis of prosopagnosia: An fMRI study. *Human Brain Mapping*, *16*(3), 176-182.

Hafri, A., Trueswell, J. C., & Epstein, R. A. (2017). Neural representations of observed actions generalize across static and dynamic visual input. *Journal of Neuroscience*, *37*(11), 3056-3071.

Hamilton, A. F. D. C., & Grafton, S. T. (2006). Goal representation in human anterior intraparietal sulcus. *Journal of Neuroscience*, *26*(4), 1133-1137.

Handwerker, D. A., Ollinger, J. M., & D'Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *NeuroImage*, *21*(4), 1639-1651.

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

Hebart, M. N., Görgen, K., & Haynes, J. D. (2015). The Decoding Toolbox (TDT): A versatile software package for multivariate analyses of functional imaging data. *Frontiers in Neuroinformatics*, *8*, 88.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, *57*(2), 243-259.

Henderson, A. M., & Woodward, A. L. (2011). "Let's work together": What do infants understand about collaborative goals?. *Cognition*, *121*, 12-21.

Hirai, M., & Hiraki, K. (2005). An event-related potentials study of biological motion perception in human infants. *Cognitive Brain Research*, *22*(2), 301-304.

Hsu, C. W., & Lin, C. J. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, *13*(2), 415-425.

Huettel, S. A., Song, A. W., & McCarthy, G. (2004). *Functional magnetic resonance imaging* (Vol. 1). Sunderland, MA: Sinauer Associates.

Huis in 't Veld, E. M., & de Gelder, B. (2015). From personal fear to mass panic: The neurological basis of crowd perception. *Human Brain Mapping*, *36*(6), 2338-2351.

Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annual Review of Psychology*, *60*, 653-670.

Iacoboni, M., Lieberman, M. D., Knowlton, B. J., Molnar-Szakacs, I., Moritz, M., Throop, C. J., & Fiske, A. P. (2004). Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline. *NeuroImage*, *21*(3), 1167-1173.

Isik, L., Koldewyn, K., Beeler, D., & Kanwisher, N. (2017). Perceiving social interactions in the posterior superior temporal sulcus. *Proceedings of the National Academy of Sciences*, *114*(43), 9145-9152.

Jacoby, N., Bruneau, E., Koster-Hale, J., & Saxe, R. (2016). Localizing Pain Matrix and Theory of Mind networks with both verbal and non-verbal stimuli. *NeuroImage*, 126, 39-48.

Jellema, T., & Perrett, D. I. (2003). Cells in monkey STS responsive to articulated body motions and consequent static posture: A case of implied motion?. *Neuropsychologia*, *41*(13), 1728-1737.

Julian, J. B., Fedorenko, E., Webster, J., & Kanwisher, N. (2012). An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *NeuroImage*, *60*(4), 2357-2364.

Kaiser, D., & Peelen, M. V. (2018). Transformation from independent to integrative coding of multi-object arrangements in human visual cortex. *NeuroImage*, *169*, 334-341.

Kana, R. K., Keller, T. A., Cherkassky, V. L., Minshew, N. J., & Just, M. A. (2009). Atypical frontal-posterior synchronization of Theory of Mind regions in autism during mental state attribution. *Social Neuroscience*, *4*(2), 135-152.

Kang, H. C., Burgund, E. D., Lugar, H. M., Petersen, S. E., & Schlaggar, B. L. (2003). Comparison of functional activation foci in children and adults using a common stereotactic space. *NeuroImage*, *19*(1), 16-28.

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

Kret, M. E., Pichon, S., Grèzes, J., & de Gelder, B. (2011). Similarities and differences in perceiving threat from dynamic faces and bodies. An fMRI study. *NeuroImage*, *54*(2), 1755-1762.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences*, *103*(10), 3863-3868.

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience*, *12*(5), 535-540.

Kriegstein, K. V., & Giraud, A. L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *NeuroImage*, *22*(2), 948-955.

Kujala, M. V., Carlson, S., & Hari, R. (2012a). Engagement of amygdala in third-person view of face-to-face interaction. *Human Brain Mapping*, *33*(8), 1753-1762.

Kujala, M. V., Kujala, J., Carlson, S., & Hari, R. (2012b). Dog experts' brains distinguish socially relevant body postures similarly in dogs and humans. *PLoS One*, *7*(6), e39145.

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., ... & Turner, R. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences*, *89*(12), 5675-5679.

Lahnakoski, J. M., Glerean, E., Salmi, J., Jääskeläinen, I. P., Sams, M., Hari, R., & Nummenmaa, L. (2012). Naturalistic FMRI mapping reveals superior temporal sulcus as the hub for the distributed brain network for social perception. *Frontiers in Human Neuroscience*, *6*, 233.

Le, T. H., Pardo, J. V., & Hu, X. (1998). 4 T-fMRI study of nonspatial shifting of selective attention: Cerebellar and parietal contributions. *Journal of Neurophysiology*, *79*(3), 1535-1548.

Lee, S. M., Gao, T., & McCarthy, G. (2012). Attributing intentions to random motion engages the posterior superior temporal sulcus. *Social Cognitive and Affective Neuroscience*, *9*, 81-87.

Lingnau, A., & Downing, P. E. (2015). The lateral occipitotemporal cortex in action. *Trends in Cognitive Sciences*, *19*(5), 268-277.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*(6843), 150-157.

Logothetis, N. K., & Wandell, B. A. (2004). Interpreting the BOLD signal. *Annual Review of Physiology*, *66*, 735-769.

Manera, V., Schouten, B., Becchio, C., Bara, B. G., & Verfaillie, K. (2010). Inferring intentions from biological motion: A stimulus set of point-light communicative interactions. *Behavior Research Methods*, *42*(1), 168-178.

Mars, R. B., Sallet, J., Schüffelgen, U., Jbabdi, S., Toni, I., & Rushworth, M. F. (2011). Connectivity-based subdivisions of the human right "temporoparietal junction area": Evidence for different areas participating in different cortical networks. *Cerebral Cortex*, *22*(8), 1894-1903.

Martin, A., & Weisberg, J. (2003). Neural foundations for understanding social and mechanical concepts. *Cognitive Neuropsychology*, *20*(3), 575-587.

Miezin, F. M., Maccotta, L., Ollinger, J. M., Petersen, S. E., & Buckner, R. L. (2000). Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *NeuroImage*, *11*(6), 735-759.

Mills, K. L., Lalonde, F., Clasen, L. S., Giedd, J. N., & Blakemore, S. J. (2012). Developmental changes in the structure of the social brain in late childhood and adolescence. *Social Cognitive and Affective Neuroscience*, *9*, 123-131.

Misaki, M., Kim, Y., Bandettini, P. A., & Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *NeuroImage*, *53*, 103-118.

Mondloch, C. J., Dobson, K. S., Parsons, J., & Maurer, D. (2004). Why 8-year-olds cannot tell the difference between Steve Martin and Paul Newman: Factors contributing to the slow development of sensitivity to the spacing of facial features. *Journal of Experimental Child Psychology*, *89*(2), 159-181.

Mosconi, M. W., Mack, P. B., McCarthy, G., & Pelphrey, K. A. (2005). Taking an "intentional stance" on eye gaze shifts: A functional neuroimaging study of social perception in children. *NeuroImage*, *27*, 247-252.

Mumford, J. A. (2012). A power calculation guide for fMRI studies. *Social Cognitive and Affective Neuroscience*, *7*(6), 738-742.

Nath, A. R., & Beauchamp, M. S. (2012). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *NeuroImage*, *59*, 781-787.

Neri, P., Luu, J. Y., & Levi, D. M. (2006). Meaningful interactions can enhance visual discrimination of human agents. *Nature Neuroscience*, *9*(9), 1186.

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

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424-430.

O'toole, A. J., Natu, V., An, X., Rice, A., Ryland, J., & Phillips, P. J. (2014). The neural representation of faces and bodies in motion and at rest. *NeuroImage*, *91*, 1-11.

Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. G., Merkle, H., & Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences*, *89*(13), 5951-5955.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97-113.

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: Multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. *Frontiers in Neuroinformatics*, *10*, 27.

Oosterhof, N. N., Tipper, S. P., & Downing, P. E. (2012). Viewpoint (in)dependence of action representations: An MVPA study. *Journal of Cognitive Neuroscience*, *24*(4), 975-989.

Osaka, N., Ikeda, T., & Osaka, M. (2012). Effect of intentional bias on agency attribution of animated motion: An event-related fMRI study. *PLoS One*, *7*(11), e49053.

Papeo, L., Corradi-Dell'Acqua, C., & Rumiati, R. I. (2011). “She” is not like “I”: The tie between language and action is in our imagination. *Journal of Cognitive Neuroscience*, *23*(12), 3939-3948.

Papeo, L., & Lingnau, A. (2015). First-person and third-person verbs in visual motion-perception regions. *Brain and Language*, *141*, 135-141.

Papeo, L., Stein, T., & Soto-Faraco, S. (2017). The two-body inversion effect. *Psychological Science*, *28*(3), 369-379.

Parkinson, C., Liu, S., & Wheatley, T. (2014). A common cortical metric for spatial, temporal, and social distance. *Journal of Neuroscience*, *34*(5), 1979-1987.

Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nature Reviews Neuroscience*, *8*(8), 636-648.

Peelen, M. V., Glaser, B., Vuilleumier, P., & Eliez, S. (2009). Differential development of selectivity for faces and bodies in the fusiform gyrus. *Developmental Science*, *12*(6), 16-25.

Pelphrey, K. A., Mitchell, T. V., McKeown, M. J., Goldstein, J., Allison, T., & McCarthy, G. (2003). Brain activity evoked by the perception of human walking: Controlling for meaningful coherent motion. *Journal of Neuroscience*, *23*(17), 6819-6825.

Pelphrey, K. A., Morris, J. P., & McCarthy, G. (2004). Grasping the intentions of others: The perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *Journal of Cognitive Neuroscience*, *16*(10), 1706-1716.

Pelphrey, K. A., Morris, J. P., Michelich, C. R., Allison, T., & McCarthy, G. (2005). Functional anatomy of biological motion perception in posterior temporal cortex: An fMRI study of eye, mouth and hand movements. *Cerebral Cortex*, *15*(12), 1866-1876.

Pelphrey, K. A., Singerman, J. D., Allison, T., & McCarthy, G. (2003). Brain activation evoked by perception of gaze shifts: The influence of context. *Neuropsychologia*, *41*(2), 156-170.

Perrett, D. I., Harries, M. H., Bevan, R., Thomas, S., Benson, P. J., Mistlin, A. J., ... & Ortega, J. E. (1989). Frameworks of analysis for the neural representation of animate objects and actions. *Journal of Experimental Biology*, *146*, 87-113.

Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., & Jeeves, M. A. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings for the Royal Society of London B, Biological Sciences, 223*(1232), 293-317.

Petrini, K., Piwek, L., Crabbe, F., Pollick, F. E., & Garrod, S. (2014). Look at those two!: The precuneus role in unattended third-person perspective of social interactions. *Human Brain Mapping*, *35*(10), 5190-5203.

Peuskens, H., Vanrie, J., Verfaillie, K., & Orban, G. A. (2005). Specificity of regions processing biological motion. *European Journal of Neuroscience*, *21*(10), 2864-2875.

Pfeiffer, U. J., Timmermans, B., Vogeley, K., Frith, C., & Schilbach, L. (2013). Towards a neuroscience of social interaction. *Frontiers in Human Neuroscience*, *7*, 22.

Pitcher, D., Dilks, D. D., Saxe, R. R., Triantafyllou, C., & Kanwisher, N. (2011). Differential selectivity for dynamic versus static information in face selective cortical regions. *NeuroImage*, *56*(4), 2356-2363.

Pitcher, D., Duchaine, B., & Walsh, V. (2014). Combined TMS and fMRI reveal dissociable cortical pathways for dynamic and static face perception. *Current Biology*, *24*(17), 2066-2070.

Quadflieg, S., Gentile, F., & Rossion, B. (2015). The neural basis of perceiving person interactions. *Cortex*, *70*, 5-20.

Quadflieg, S., & Koldewyn, K. (2017). The neuroscience of people watching: How the human brain makes sense of other people's encounters. *Annals of the New York Academy of Sciences*, *1396*, 166-182.

Quinn, P. C., Yahr, J., Kuhn, A., Slater, A. M., & Pascalis, O. (2002). Representation of the gender of human faces by infants: A preference for female. *Perception*, *31*(9), 1109-1121.

Redcay, E. (2008). The superior temporal sulcus performs a common function for social and speech perception: Implications for the emergence of autism. *Neuroscience & Biobehavioral Reviews*, *32*, 123-142.

Redcay, E., Dodell-Feder, D., Pearrow, M. J., Mavros, P. L., Kleiner, M., Gabrieli, J. D., & Saxe, R. (2010). Live face-to-face interaction during fMRI: A new tool for social cognitive neuroscience. *NeuroImage*, *50*(4), 1639-1647.

Redcay, E., Velnoskey, K. R., & Rowe, M. L. (2016). Perceived communicative intent in gesture and language modulates the superior temporal sulcus. *Human Brain Mapping*, *37*(10), 3444-3461.

Reed, C. L., Stone, V. E., Bozova, S., & Tanaka, J. (2003). The body-inversion effect. *Psychological Science*, *14*(4), 302-308.

Richardson, H., Lisandrelli, G., Riobueno-Naylor, A., & Saxe, R. (2018). Development of the social brain from age three to twelve years. *Nature Communications*, *9*, 1027.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*(2), 131-141.

Roberts, K. L., & Humphreys, G. W. (2010). Action relationships concatenate representations of separate objects in the ventral visual system. *NeuroImage*, *52*(4), 1541-1548.

Röder, B., Stock, O., Neville, H., Bien, S., & Rösler, F. (2002). Brain activation modulated by the comprehension of normal and pseudo-word sentences of different processing demands: A functional magnetic resonance imaging study. *NeuroImage*, *15*(4), 1003-1014.

Ross, P. D., de Gelder, B., Crabbe, F., & Grosbras, M. H. (2014). Body selective areas in the visual cortex are less active in children than in adults. *Frontiers in Human Neuroscience*, *8*, 941.

Roux, P., Passerieux, C., & Ramus, F. (2013). Kinematics matters: A new eye-tracking investigation of animated triangles. *The Quarterly Journal of Experimental Psychology*, *66*(2), 229-244.

Saggar, M., Shelly, E. W., Lepage, J. F., Hoeft, F., & Reiss, A. L. (2014). Revealing the neural networks associated with processing of natural social interaction and the related effects of actor-orientation and face-visibility. *NeuroImage*, *84*, 648-656.

Sapey-Triomphe, L. A., Centelles, L., Roth, M., Fonlupt, P., Henaff, M. A., Schmitz, C., & Assaiante, C. (2017). Deciphering human motion to discriminate social interactions: A developmental neuroimaging study. *Social Cognitive and Affective Neuroscience*, *12*(2), 340-351.

Saxe, R., Brett, M., & Kanwisher, N. (2006). Divide and conquer: A defense of functional localizers. *NeuroImage*, *30*(4), 1088-1096.

Saxe, R., Jamal, N., & Powell, L. (2005). My body or yours? The effect of visual perspective on cortical body representations. *Cerebral Cortex*, *16*(2), 178-182.

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. *Neuropsychologia*, *42*(11), 1435-1446.

Scherf, K. S., Behrmann, M., Humphreys, K., & Luna, B. (2007). Visual category-selectivity for faces, places and objects emerges along different developmental trajectories. *Developmental Science*, *10*(4), F15-F30.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience 1. *Behavioral and Brain Sciences*, *36*(4), 393-414.

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences*, *4*(8), 299-309.

Schultz, J., Friston, K. J., O'Doherty, J., Wolpert, D. M., & Frith, C. D. (2005). Activation in posterior superior temporal sulcus parallels parameter inducing the percept of animacy. *Neuron*, *45*(4), 625-635.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews*, *42*, 9-34.

Shih, P., Keehn, B., Oram, J. K., Leyden, K. M., Keown, C. L., & Müller, R. A. (2011). Functional differentiation of posterior superior temporal sulcus in autism: A functional connectivity magnetic resonance imaging study. *Biological Psychiatry*, *70*(3), 270-277.

Shultz, S., Lee, S. M., Pelphrey, K., & McCarthy, G. (2010). The posterior superior temporal sulcus is sensitive to the outcome of human and non-human goal-directed actions. *Social Cognitive and Affective Neuroscience*, 6(5), 602-611.

Shultz, S., Vouloumanos, A., & Pelphrey, K. (2012). The superior temporal sulcus differentiates communicative and noncommunicative auditory signals. *Journal of Cognitive Neuroscience*, *24*(5), 1224-1232.

Sinke, C. B., Sorger, B., Goebel, R., & de Gelder, B. (2010). Tease or threat? Judging social interactions from bodily expressions. *NeuroImage*, *49*(2), 1717-1727.

Sladky, R., Friston, K. J., Tröstl, J., Cunnington, R., Moser, E., & Windischberger, C. (2011). Slice-timing effects and their correction in functional MRI. *NeuroImage*, *58*(2), 588-594.

Slater, A., Quinn, P. C., Hayes, R., & Brown, E. (2000). The role of facial orientation in newborn infants' preference for attractive faces. *Developmental Science*, *3*(2), 181-185.

Sliwa, J., & Freiwald, W. A. (2017). A dedicated network for social interaction processing in the primate brain. *Science*, *356*(6339), 745-749.

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*(1), 83-98.

Soares, J. M., Magalhães, R., Moreira, P. S., Sousa, A., Ganz, E., Sampaio, A., ... & Sousa, N. (2016). A hitchhiker's guide to functional magnetic resonance imaging. *Frontiers in Neuroscience*, *10*, 515.

Spunt, R. P., & Adolphs, R. (2014). Validating the why/how contrast for functional MRI studies of theory of mind. *NeuroImage*, *99*, 301-311.

Strother, S. C. (2006). Evaluating fMRI preprocessing pipelines. *IEEE Engineering in Medicine and Biology Magazine*, *25*(2), 27-41.

Szego, P. A., & Rutherford, M. D. (2007). Actual and illusory differences in constant speed influence the perception of animacy similarly. *Journal of Vision*, *7*(12), 5.

Tavares, P., Lawrence, A. D., & Barnard, P. J. (2007). Paying attention to social meaning: An FMRI study. *Cerebral Cortex*, *18*(8), 1876-1885.

Thurman, S. M., & Lu, H. (2014). Perception of social interactions for spatially scrambled biological motion. *PLoS One*, *9*(11), e112539.

Trapp, K., Spengler, S., Wüstenberg, T., Wiers, C. E., Busch, N. A., & Bermpohl, F. (2014). Imagining triadic interactions simultaneously activates mirror and mentalizing systems. *NeuroImage*, *98*, 314-323.

Tsoi, L., Dungan, J., Waytz, A., & Young, L. (2016). Distinct neural patterns of social cognition for cooperation versus competition. *NeuroImage*, *137*, 86-96.

Tucciarelli, R., Turella, L., Oosterhof, N. N., Weisz, N., & Lingnau, A. (2015). MEG multivariate analysis reveals early abstract action representations in the lateral occipitotemporal cortex. *Journal of Neuroscience*, *35*(49), 16034-16045.

Van den Stock, J., Hortensius, R., Sinke, C., Goebel, R., & De Gelder, B. (2015). Personality traits predict brain activation and connectivity when witnessing a violent conflict. *Scientific Reports*, *5*, 13779.

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

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *NeuroImage*, *48*(3), 564-584.

Vander Wyk, B. C., Hudac, C. M., Carter, E. J., Sobel, D. M., & Pelphrey, K. A. (2009). Action understanding in the superior temporal sulcus region. *Psychological Science*, *20*(6), 771-777.

Vander Wyk, B. C., Voos, A., & Pelphrey, K. A. (2012). Action representation in the superior temporal sulcus in children and adults: An fMRI study. *Developmental Cognitive Neuroscience*, *2*(4), 409-416.

Vangeneugden, J., Peelen, M. V., Tadin, D., & Battelli, L. (2014). Distinct neural mechanisms for body form and body motion discriminations. *Journal of Neuroscience*, *34*(2), 574-585.

Vanrie, J., & Verfaillie, K. (2004). Perception of biological motion: A stimulus set of human point-light actions. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 625-629.

Vestner, T., Tipper, S. P., Hartley, T., Over, H., & Rueschemeyer, S. A. (2018). Bound Together: Social binding leads to faster processing, spatial distortion and enhanced memory of interacting partners. *Journal of Experimental Psychology: General* (in press).

Wagner, D. D., Kelley, W. M., Haxby, J. V., & Heatherton, T. F. (2016). The dorsal medial prefrontal cortex responds preferentially to social interactions during natural viewing. *Journal of Neuroscience*, *36*(26), 6917-6925.

Walbrin, J., Downing, P., & Koldewyn, K. (2018). Neural responses to visually observed social interactions. *Neuropsychologia*, *112*, 31-39.

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*(10), 1854-1863.

Wang, Y., & Quadflieg, S. (2015). In our own image? Emotional and neural processing differences when observing human–human vs human–robot interactions. *Social Cognitive and Affective Neuroscience*, *10*(11), 1515-1524.

Wende, K. C., Nagels, A., Blos, J., Stratmann, M., Chatterjee, A., Kircher, T., & Straube, B. (2013). Differences and commonalities in the judgment of causality in physical and social contexts: An fMRI study. *Neuropsychologia*, *51*(13), 2572-2580.

Wise, R. J., Scott, S. K., Blank, S. C., Mummery, C. J., Murphy, K., & Warburton, E. A. (2001). Separate neural subsystems within Wernicke's area'. *Brain*, *124*, 83-95.

Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., & Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, *4*, 58-73.

Wurm, M. F., Caramazza, A., & Lingnau, A. (2017). Action categories in lateral occipitotemporal cortex are organized along sociality and transitivity. *Journal of Neuroscience*, *37*(3), 562-575.

Wurm, M. F., & Schubotz, R. I. (2018). The role of the temporoparietal junction (TPJ) in action observation: Agent detection rather than visuospatial transformation. *NeuroImage*, *165*, 48-55.

Wurm, M. F., von Cramon, D. Y., & Schubotz, R. I. (2011). Do we mind other minds when we mind other minds' actions? A functional magnetic resonance imaging study. *Human Brain Mapping*, *32*(12), 2141-2150.

Wutte, M. G., Smith, M. T., Flanagin, V. L., & Wolbers, T. (2011). Physiological signal variability in hMT+ reflects performance on a direction discrimination task. *Frontiers in Psychology*, *2*, 185.

Xiong, J., Gao, J. H., Lancaster, J. L., & Fox, P. T. (1995). Clustered pixels analysis for functional MRI activation studies of the human brain. *Human Brain Mapping*, *3*(4), 287-301.

Yang, D. Y. J., Rosenblau, G., Keifer, C., & Pelphrey, K. A. (2015). An integrative neural model of social perception, action observation, and theory of mind. *Neuroscience & Biobehavioral Reviews*, *51*, 263-275.

Yang, J., Andric, M., & Mathew, M. M. (2015). The neural basis of hand gesture comprehension: A meta-analysis of functional magnetic resonance imaging studies. *Neuroscience & Biobehavioral Reviews*, *57*, 88-104.

Yarkoni, T. (2009). Big correlations in little studies: Inflated fMRI correlations reflect low statistical power—Commentary on Vul et al.(2009). *Perspectives on Psychological Science*, *4*(3), 294-298.

Zieber, N., Kangas, A., Hock, A., & Bhatt, R. S. (2014). Infants' perception of emotion from body movements. *Child Development*, *85*(2), 675-684.

Zumer, J. M., Brookes, M. J., Stevenson, C. M., Francis, S. T., & Morris, P. G. (2010). Relating BOLD fMRI and neural oscillations through convolution and optimal linear weighting. *NeuroImage*, *49*(2), 1479-1489.

Appendices

Chapters 3, 4, & 5

Chapter 3 Appendix

A. Pilot Study

8 participants were recruited from the Bangor University population (4 females; aged 19 – 29, *M* = 25.2 years, *SD* = 3.55). The stimuli, design, interaction localizer, and fMRI acquisition parameters were almost identical to those used in experiment 2, except that competitive interactions always resulted in failed goal-outcome (e.g. a ball was not successfully pushed into a goal), and cooperative interactions resulted in successful goal outcome (e.g. a ball was successfully pushed into a goal). Participants were scanned whilst watching moving shape videos that depicted either interactive (i.e. competitive or cooperative) or non-interactive scenarios. Percent signal change was extracted from pSTS ROIs for each condition, and entered into t-tests. Significant differences were found in both the interaction > non-interaction contrast ($t(7) = 3.16$, $p=.008$), and the competition > cooperation contrast ($t(7) = -1.94$, $p=.047$), providing preliminary evidence for pSTS sensitivity to moving shape interactions. Additionally, univariate whole brain analysis revealed strong activation in the right LOTC for the interaction > non-interaction contrast. As this strong peak differed from the experiment 1 results, we included a 6mm sphere centered on the peak of this response as a region of interest in experiment 2.

B. ROI Lateralization

We observed some variability in whole brain peak activation for the interaction localizer task: 10 participants demonstrated peak global activation peak in the right pSTS, whilst the left pSTS contained peak clusters in 6 participants, and non-pSTS peaks were observed in 5 participants. The decision to restrict pSTS ROI analyses to the right hemisphere was motivated by several factors: firstly, that right pSTS activity was observed across all but 2 participants – those with left sided peaks tended to show right activity too; secondly, comparison of classification performance in left lateralized participants – for both left and right pSTS ROIs – demonstrated equivalent classification performance in both ROIs; and thirdly, previous interaction perception studies tend to show right lateralized pSTS activation maxima at group level. For the TPJ localizer task, most participants showed right-lateralized peak activation with the right TPJ area, and so right TPJ ROIs were selected.

C. ROI Size and Location

4 participants had slightly overlapping pSTS and TPJ ROIs (i.e. mean overlap = 9.25 voxels), and so overlapping voxels were removed from the final ROIs. To determine if there was reliable spatial separation between pSTS and TPJ ROIs (see supplementary figure 1), we calculated three paired t-tests (two-tailed) based on the central MNI coordinate in each ROI, for each of the x, y, and z dimensions. The two ROIs did not differ in lateral placement (x-dimension: $t(14) = 1.69$, $p=.113$), but pSTS ROIs were significantly more anterior (y-dimension: $t(14) = 4.10$, $p=.001$) and ventral (z-dimension: $t(14) = -2.35$, $p=.034$) than TPJ ROIs. A paired t-test for the number of voxels in pSTS ($M = 102.53$, $SD = 15.16$) and TPJ ROIs ($M = 108.60$, $SD = 14.96$), revealed no difference ($t(14) = -1.08$, $p=.298$). In comparison, the LOTC ROI contained 123 voxels.

D. Motion Energy Analysis

We conducted a 2 x 2 factorial ANOVA (i.e. interaction and non-interaction as levels of the first factor, with competition and cooperation as levels of the second factor) for stimulus motion energy values for each condition. Neither main effect between interaction and non-interaction ($F(1,60) = 0.55$, p=.462) or main effect of competition and cooperation ($F(1,60) = 0.03$, p=.866), nor interaction term ($F(1,60) = 0.05$, p=.826) was significant, indicating that motion energy did not differ between conditions.

E. Stimulus Ratings

An independent group of participants (N=20) Likert-rated (i.e. 1 = strongly disagree, 7 = strongly agree) a subset of stimulus videos, and these ratings were entered into three 2x2 ANOVAs. For the question 'The agents interacted with each other', a main effect of interaction and non-interaction was observed ($F(1,19) = 180.96$. $p<.001$), but no main effect between competition and cooperation ($F(1,19) = 0.86$. $p=.366$), nor interaction term was significant($F(1,17) = 0.00$. $p=.975$). For the question 'The agents were goal-directed', a main effect of interaction and non-interaction was observed ($F(1,19) = 115.13$. $p<.001$), but no main effect between competition and cooperation ($F(1,19) = 2.98$. $p=.100$), nor interaction term was significant($F(1,17) = 2.40$. $p=.138$). For the question 'The agents were alive/animate', a main effect of interaction and non-interaction was observed ($F(1,19) = 46.52$. $p<.001$), but no main effect between competition and cooperation ($F(1,19) = 0.94$. p=.762), nor interaction term was significant ($F(1,17) = 0.42$. $p=.524$).

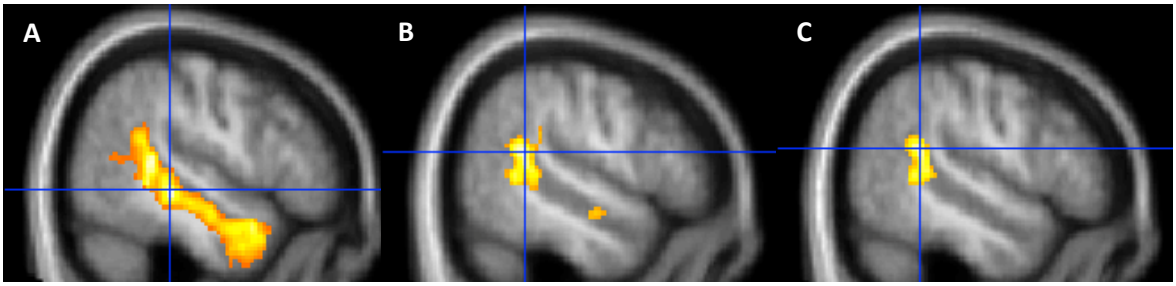
F. Example Moving Shape Stimuli

Example moving shape stimuli can be viewed via following link:

https://drive.google.com/open?id=1sjcgU9Wo0XN90MLr_ZrOwmEAdcxNE8PJ

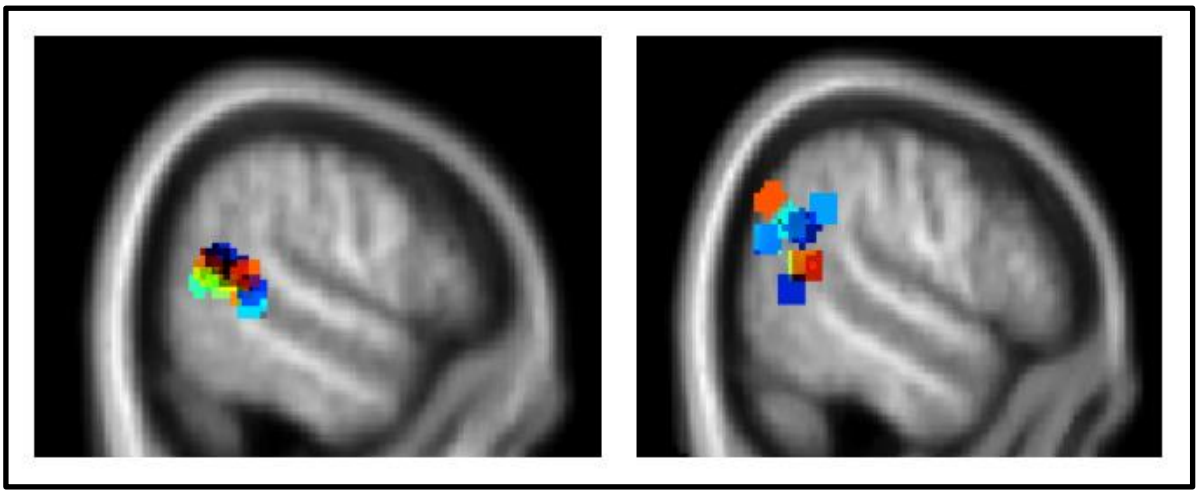
G. Supplementary Figures

Supplementary Figure 1



Supplementary figure 1. Whole brain activation maps for the interaction localizer task (i.e. point-light human interactions > individual actions). Height threshold = .001(uncorrected), FDR cluster correction (p<.05). A) Localizer data (used to localise pSTS ROIs) in experiment 2 (N=19; peak whole brain activation is shown at MNI coordinates (x,y,z) 50 -32 -4; B) Localizer data from an independent participant group (not included in either experiment 1 or 2; N=20; peak whole brain activation: 54 -44 16; C) Conjunction analysis between the two datasets in A and B was performed in SPM12 and revealed a single right pSTS cluster (peak MNI coordinates: 52 -44 18; (height threshold = .001(uncorrected), FDR cluster correction (p<.05).

Supplementary Figure 2



Supplementary figure 2. Subject-specific localized ROIs. Left: Right pSTS. Right: Right TPJ.

Table A

Mean centre of ROI coordinates

Dimension	pSTS	TPJ
X	57.23 (5.90)	53.60 (6.01)
Y	-46.04 (6.02)	-54.88 (6.55)
Z	13.26 (10.39)	20.51 (8.40)

Note. Values in parentheses are SD.

Table B

Run omissions across participants

Participant	Number of Runs Excluded from Analyses	Runs	Reason for Omission
1	1	4	Presentation Script Error
2	1	4	Presentation Script Error

3	1	4	Presentation Script Error
4	1	4	Presentation Script Error
5	2	4,10	Presentation Script Error; <75% response accuracy
6	1	4	Presentation Script Error
7	1	5	>0.5mm Motion Spikes
13	3	1,2,9	>0.5mm Motion Spikes (runs 1 and 2); <75% response accuracy
16	1	9	>0.5mm Motion Spikes
17	1	1	>0.5mm Motion Spikes
20	1	8	>0.5mm Motion Spikes
25	1	9	>0.5mm Motion Spikes

Table C

Participant omissions from analyses

Participant	Analyses Excluded From	Reason for Omission
1	TPJ ROI analyses	Pilot participant – no TPJ localization
2	TPJ ROI analyses	Pilot participant – no TPJ localization
3	TPJ ROI analyses	Pilot participant – no TPJ localization
4	ALL	No button responses
9	pSTS ROI analyses	No localizable clusters
10	TPJ ROI analyses	No localizable clusters
11	ALL	< 50% response accuracy across numerous runs
17	Supplementary Searchlight Analyses	Frontal lobe signal drop-out due to dental brace
18	pSTS ROI analyses	No localizable clusters

Chapter 4 Appendix

A. *Search Space Size Descriptive Statistics*

	Adults		Children	
	M	SD	M	SD
rpSTS-I	229.00	0.00	273.23	1.26
lpSTS-I	229.00	0.00	280.00	0.00
rSTS-F	250.75	1.65	265.47	1.50
lSTS-F	234.46	4.88	238.23	4.93
rEBA	223.29	0.60	209.07	1.02
lEBA	228.89	3.46	241.90	6.49
rTPJ-M	229.00	0.00	262.90	2.24
lTPJ-M	217.24	0.63	268.00	0.00
rFFA	156.82	2.47	160.90	3.55

Means and standard deviations for voxels are shown.

B. *Initial Head Motion Analysis*

Head motion was first compared between age groups, and localizer tasks using the average root mean square (RMS) measure as described in the methods section. A 3 x 2 (localizer task x age group) mixed ANOVA revealed a main effect of age group ($F(1,56) = 30.24$, $p < .001$, $\eta_p^2 = .240$) demonstrating that head motion was greater in children than adults, as is previously reported (e.g. Kang, Burgund, Lugar, Peterson, & Schlaggar, 2003). A main effect of task ($F(1.41,78.85) = 17.71$, $p < .001$, $\eta_p^2 = .351$), and interaction between factors was also observed ($F(1.41,78.85) = 11.01$, $p < .001$, $\eta_p^2 = .165$).

C. *Condition-wise Comparisons of Head Motion*

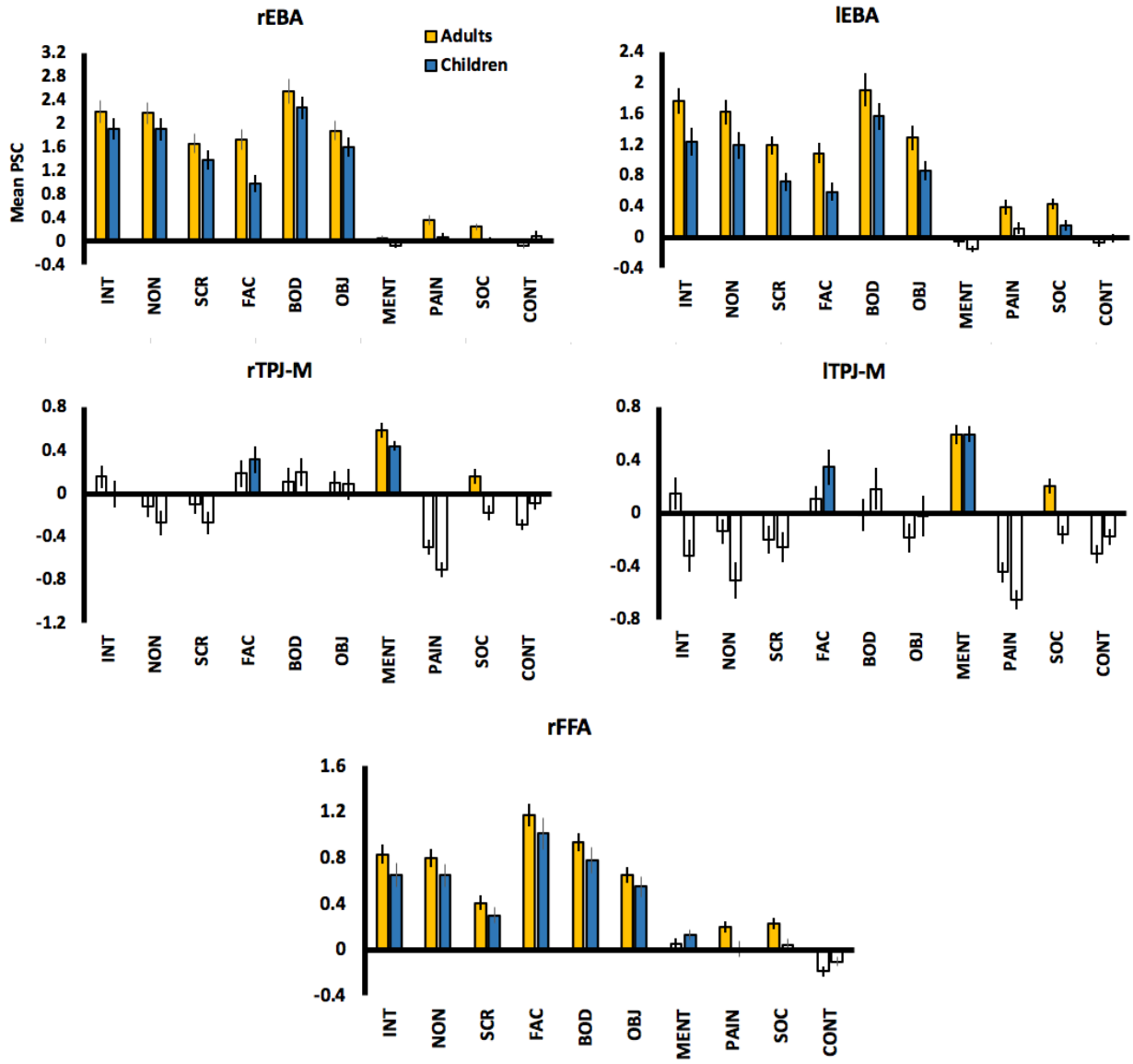
RMS head motion comparisons were performed with a three 2 x 2 mixed ANOVAs (selectivity contrast pair x age group) to determine potential group differences for each selectivity measure. For the faces > objects contrast, a main effect of group ($F(1,56) = 12.66$, $p = .001$, $\eta_p^2 = .184$) but neither main effect of condition nor interaction reached significance (both $ps > .696$). Similarly, for the bodies > objects contrast, a main effect of group ($F(1,56) = 12.24$, $p = .001$, $\eta_p^2 = .179$) but neither main effect of condition nor interaction reached significance (both $ps > .415$). Therefore, despite greater head movement for children in these two contrasts, this did not differ between these

conditions, and therefore head motion cannot account for any selectivity differences for these contrasts. However, for the mentalizing > pain contrast, a main effect of condition ($F(1,58) = 4.55$, $p = .037$, $\eta_p^2 = .073$), but neither main effect of group nor interaction was shown (both *ps* > .207).

D. Initial PSC Analyses: Hemisphere x Condition pSTS-I ANOVA

To determine whether PSC for the interaction condition was greater in the right than left pSTS-I, a 2x2 mixed ANOVA (hemisphere x age group) was performed. This revealed a main effect of hemisphere ($F(1,58) = 6.05$, $p = .017$, $\eta_p^2 = .094$), no main effect of group ($F(1,58) = 0.27$, $p = .607$, $\eta_p^2 = .005$), or interaction ($F(1,58) = 0.81$, $p = .373$, $\eta_p^2 = .014$), demonstrating stronger responses in the right pSTS, in both groups.

E. PSC Charts for Bilateral EBA, TPJ-M, & right FFA



F. Right pSTS-I is the Most Interaction Selective ROI: Full Statistics

To determine if right pSTS-I selectivity was greater than for other regions (right EBA, left EBA, & right FFA) ROIs, 2 x 2 mixed ANOVAs (region x age group) were performed. For the right EBA, a main effect of region ($F(1,55) = 36.09$, $p < .001$, $\eta_p^2 = .396$), no main effect of group ($F(1,55) = 1.78$, $p = .187$, $\eta_p^2 = .031$), and marginal interaction ($F(1,55) = 3.18$, $p = .080$, $\eta_p^2 = .055$) was shown. For the left EBA, a main effect of region ($F(1,55) = 16.39$, $p < .001$, $\eta_p^2 = .230$), no main effect of group ($F(1,55) = 2.98$, $p = .090$, $\eta_p^2 = .051$), and no interaction ($F(1,55) = 0.86$, $p = .357$, $\eta_p^2 = .015$) was observed. And for the right FFA, a main effect of region ($F(1,55) = 36.43$, $p < .001$, $\eta_p^2 = .398$), no main effect of group ($F(1,55) = 1.96$, $p = .168$, $\eta_p^2 = .034$), and marginal interaction ($F(1,55) = 3.19$, $p = .080$, $\eta_p^2 = .055$). Therefore, right pSTS-I selectivity was greater than for all three regions, with no main effect of group, and marginal interaction trends for the right EBA and right FFA comparison only.

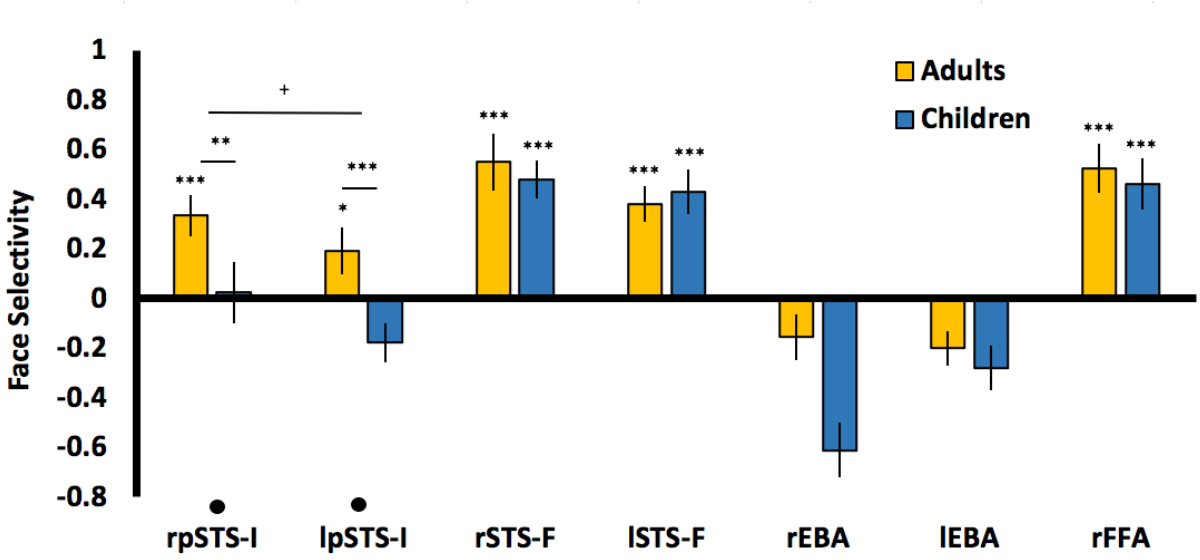
Follow-up paired t-tests (one tailed) were performed due to marginal interaction terms in the right pSTS-I vs. other regions ANOVAs, and revealed the following significant effects: right pSTS > right EBA in adults ($t(27) = 4.71$, $p < .001$); for right pSTS-I > right EBA in children ($t(28) = 3.69$, $p = .001$); right pSTS-I > right FFA in adults ($t(27) = 5.85$, $p < .001$); right pSTS-I > right FFA in children ($t(28) = 2.87$, $p = .004$, one-tailed). Because adults – but not children – showed above-zero selectivity in the bilateral STS-F ROIs, we ran two paired t-tests for adults only. Right pSTS-I selectivity was significantly greater than both the right ($t(27) = 2.61$, $p = .008$, one-tailed) and left STS-F ($t(27) = 3.30$, $p = .002$, one-tailed). Therefore, interaction selectivity was significantly greater in the right pSTS-I than all other ROIs, for both groups.

To compare selectivity in the left pSTS-I with all other regions (except right pSTS-I), in adults (but not children as they did not show above zero selectivity in the left pSTS-I), paired t-tests were performed. These revealed that interaction selectivity was greater in the left pSTS-I was greater than the right FFA ($t(27) = 3.86$, $p = .001$, one-tailed) and right EBA ($t(27) = 3.05$, $p = .003$, one-tailed), and greater than the left EBA ($t(27) = 2.01$, $p = .027$, one-tailed) and left STS-F ($t(27) = 1.77$, $p = .044$, one-tailed) at an uncorrected threshold (i.e. $p < .05$), whilst the trend towards greater selectivity than the right STS-F was not significant ($t(27) = 3.05$, $p = .003$, one-tailed). These results demonstrate a moderate trend for greater left pSTS-I selectivity compared to other regions.

G. Face Selectivity Analyses

Given the observation for relatively strong interaction selectivity in STS-F, and the close proximity of this region to the pSTS-I – face selective responses were examined. First, face selective responses between right and left pSTS-I ROIs were tested with a 2 x 2 mixed ANOVA (region x age group). A marginal main effect of region ($F(1,55) = 3.76$, $p = .057$, $\eta_p^2 = .064$), main effect of age group ($F(1,55) = 10.36$, $p = .002$, $\eta_p^2 = .158$) and no interaction was observed ($F(1,55) = 0.11$, $p = .737$, $\eta_p^2 = .002$). These findings show that face selectivity is greater for adults than children in bilateral pSTS-I. A marginal trend is also shown for greater selectivity in the right hemisphere for both groups (although

children did not show above zero selectivity in either hemisphere). A similar analysis in the bilateral STS-F revealed no significant ANOVA term, demonstrating that face selectivity is statistically the same between age groups, in both STS-F ROIs (no main effect of hemisphere: $F(1,55) = 2.23$, $p = .141$, $\eta_p^2 = .039$; no main effect of age group: $F(1,55) = 0.01$, $p = .939$, $\eta_p^2 = .000$; no interaction between factors: $F(1,55) = 0.68$, $p = .413$, $\eta_p^2 = .012$).



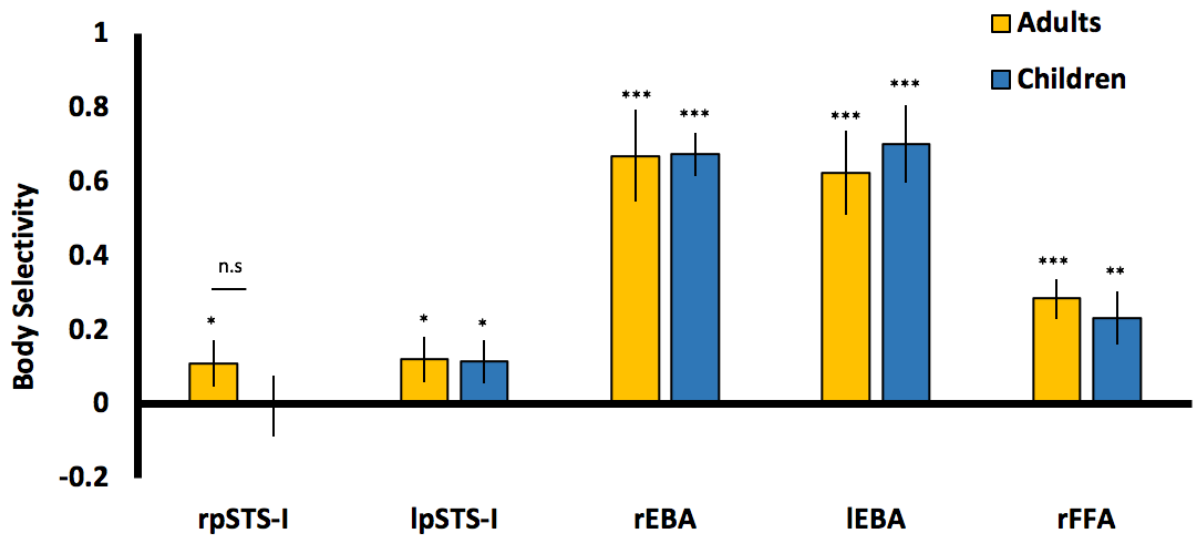
Next, to determine if face selectivity significantly differed between pSTS-I and STS-F regions, two further ANOVAs were performed. Face selectivity for the right STS-F was significantly greater than the right pSTS-I ($F(1,55) = 13.64$, $p = .001$, $\eta_p^2 = .199$), along with a marginal main effect of age group ($F(1,55) = 2.88$, $p = .095$, $\eta_p^2 = .050$), and no interaction between factors ($F(1,55) = 1.78$, $p = .187$, $\eta_p^2 = .031$). As with these right hemisphere ROIs, a significant main effect of region ($F(1,55) = 21.81$, $p < .001$, $\eta_p^2 = .284$), a marginal main effect of age group ($F(1,55) = 3.58$, $p = .064$, $\eta_p^2 = .061$) was shown between left STS-F and left pSTS-I, however, a significant interaction ($F(1,55) = 6.20$, $p = .016$, $\eta_p^2 = .101$) also emerged. Follow up t-tests demonstrate greater face selectivity for faces in the left STS-F than pSTS-I for children ($t(28) = 4.98$, $p = .001$) but only marginally in adults ($t(27) = 1.57$, $p = .064$) – an effect likely driven by 'negative selectivity' for faces in the pSTS-I in children. Therefore, face selectivity in right STS-F was significantly greater than in right pSTS-I, in both age groups, however, a similar difference was seen for children only in the left hemisphere.

		M	SD	t	df	p
rpSTS-I	Adults	0.37	0.44	4.01	27	<.001
	Children	0.02	0.67	0.20	28	.846
lpSTS-I	Adults	0.19	0.51	2.02	27	.053
	Children	-0.18	0.43	-2.23	28	n.s
rSTS-F	Adults	0.55	0.61	4.78	27	<.001
	Children	0.48	0.41	6.36	28	<.001
lSTS-F	Adults	0.38	0.38	5.28	27	<.001
	Children	0.43	0.48	4.83	28	<.001

One-sample t-tests reveal that bilateral STS-F face selectivity is significantly greater than zero in both adults and children. Adults demonstrate above zero selectivity in the right pSTS-I, and marginally in the left pSTS-I, whereas children are not above zero in either hemisphere.

How do these patterns of results relate to interaction selectivity? One general trend is observed: Adults show strongest selectivity in corresponding selective regions (e.g. interactions in pSTS-I and faces in STS-F), but also show weaker responses in the 'other' non-selective STS region (along with other regions). By contrast, children show selective responses in selective regions but not non-selective regions of the STS. Together, these findings suggest that adults recruit wider brain areas during interaction and face perception, whereas children show a greater reliance on selective cortex.

H. Body Selectivity Analyses



To test whether body selective responses were above-zero in the pSTS-I ROIs, one-sample t-tests (one tailed) were performed. Both adults (M = 0.12, SD = 0.32; $t(27)$ = 1.98, p = .029) and children (M = 0.11, SD = 0.32; $t(28)$ = 1.95, p = .031) showed weak selectivity

in the left pSTS-I. By contrast, adults (M = 0.11, SD = 0.32; $t(27)$ = 1.79, p =.043) but not children (M = 0.00, SD = 0.44; $t(28)$ = -0.06, p =.981) showed weak body selectivity in the right pSTS-I.

To compare whether this weak selectivity differed between pSTS-I ROIs, between groups, a 2 x 2 mixed ANOVA (hemisphere x age group) ANOVA was performed. Neither a main effect of hemisphere ($F(1,55)$ = 1.48, p = .228, η_p^2 = .026), main effect of age group ($F(1,55)$ = 0.60, p = .444, η_p^2 = .011), nor significant interaction ($F(1,55)$ = 1.05, p = .309, η_p^2 = .019) was found. Therefore, no group or hemisphere differences were observed for weak trends towards body selective responses in the pSTS-I ROIs. Additionally, PSC responses show that both adults and children show strong body responses in bilateral EBA and right FFA.

Chapter 5 Appendix

A. Scripted Action-gestures & Stimulus Creation Description

	Action-gesture Variant	Actor A	Actor B
Arguing	1	*With one arm, making short, sharp, 'accusatory' pointing gestures towards B.*	*Arms crossed, making short, sharp 'confrontational' head movements towards A.*
	2	*Stamping one leg, raising fists and throwing them down in frustration.*	*With both hands, making alternating short, sharp gestures towards either side of themselves (as if listing reasons for their anger).*
	3	*Leaning forward, counting on their fingers in a short, sharp, fashion, (as if listing reasons for their anger).*	*Bringing both hands up to their head and throwing them down/towards A in a frustrated manner.*
	4	*Leaning forward, gesturing sharply towards the B with their middle finger raised.*	*Leaning forward in a confrontational manner, raising arms and throwing them straight down.*
Celebrating	1	*Leaning back a little, shaking each fist in an excited fashion.*	*Leaning back a little, throwing arms out, thumbs up, celebrating.*
	2	*Throwing arms up and pumping them in unison.*	*Bending over a little, shaking fists in front of torso excitedly.*
	3	*With hands together in praying gesture, shaking them in front of head.*	*Arms up shaking them, jumping up and down a few times.*
	4	*Alternating pumping fists and clapping.*	*Leaning back and shaking arms, fingers pointed upwards.*
Laughing	1	Leaning forward pointing and laughing.	Leaning back, raising hand to head, convulsing.
	2	Holding stomach, convulsing.	Raising hand to head and gesturing towards A.
	3	Hands on legs, rocking back and forth.	Convulsing, looking up.
	4	Smacking leg, leaning forward.	Raising hand, covering mouth.

Stimuli were created from video footage of three different female-male actor pairs filmed in side profile, engaging in semi-scripted interactions in front of a green screen (allowing for the actors to be superimposed on a plain grey background). Actors were asked to perform scenarios for three qualitatively different interaction scenarios: *Arguing*, *celebrating* and *laughing*. The timing of scenarios was not explicitly controlled, but instead the actors were encouraged to perform relatively extended improvised sequences (i.e. up to 30 seconds) – this allowed the actors to produce more flowing and natural scenarios, from which we could choose the most suitable segments as the final stimuli.

For each interaction scenario, different action-gestures were always scripted for each actor, and these gestures were always congruent with the type of interaction (e.g. both performed different angry or frustrated gestures while arguing). Each of the three actor pairs performed a total of 8 arguing scenarios, 7 celebrating scenarios, and 5 laughing scenarios. To create the final stimuli, 4 second segments of footage were selected that were deemed to most convincingly demonstrated the intended interaction scenarios. A pilot rating study (N=10) was then used to determine, for each interaction scenario, the four videos that depicted unique action-gesture pairings with the highest ratings for *interactiveness* and *naturalness* (i.e. participants gave Likert responses to the following questions: 'How interactive did the interactors appear to be?'; 'How natural did the interaction appear to be?'); these were the final stimuli used in the main study.

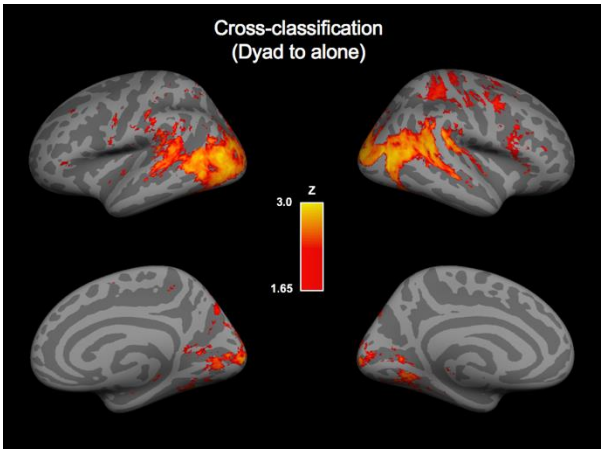
12 example video stimuli (for one of the interactor pairs) can be downloaded from the following link:

https://drive.google.com/open?id=1kEUJNiZWu00dyDc1Id8m6JZ9Hx-Ozyj0

B. One-Sample Statistics for STS-F and TPJ-M Classification

Region		**Dyad Classification**				**Alone Classification**			
	df	**M (%)**	**SD**	**t**	**p**	**M**	**SD**	**t**	**p**
rTPJ-M	17	35.19	13.34	0.59	.282	35.80	9.38	1.12	.140
lTPJ-M	16	32.68	11.77	-0.23	.589	35.95	10.60	1.02	.162
rSTS-F	15	37.50	12.42	1.34	.100	37.85	12.70	1.42	.088
lSTS-F	12	38.39	16.04	1.25	.118	36.33	9.79	1.10	.146

C. SVM Searchlight Analysis: Cross-classification



Surface-registered whole brain searchlight results for cross-classification. Colour bar represents TFCE z-scores.