Unraveling Top-Down and Bottom-Up Processes in Theory of Mind with Layer fMRI

Master's Thesis

Requirement for Master's Degree (MSc)

at the Faculty of Natural and Life Sciences

at the Paris Lodron University Salzburg

Submitted by
Leonard Elia van Dyck, BSc
Matriculation no. 11728153

Supervisor
Assoc.-Prof. Dr. Mario Braun
Co-Supervisor
Assoc.-Prof. Dr. Martin Kronbichler

Department of Psychology

Salzburg, January 2023

Abstract

Theory of Mind (ToM), the ability to reason about the mental states of others, is a crucial social skill of humans. Neuroimaging studies have found that the *Temporo-Parietal Junction* (TPJ) and an entire network of additional brain regions are specifically activated during mental state reasoning. Two common tasks that are used to investigate ToM are the *False Belief* (FB) task and the *Social Animations* (SA) task, which are thought to activate the posterior part (pTPJ) and the anterior part (aTPJ) of TPJ, respectively. While previous research has suggested potential explanations for this functional specialization, the exact mechanisms are not yet fully understood. In this study, high-resolution layer fMRI was used to examine neural activity in cortical layers of pTPJ and aTPJ during the FB and SA tasks. Firstly, the results corroborate the importance of the ToM network and especially TPJ in mental state reasoning. Secondly, a significant interaction between task and region was revealed, which underlines the expected functional specialization of TPJ clusters. Thirdly, the layer profiles of the two tasks indicated feedback-like activity, but when separated by region, pTPJ showed feedback-like activity for the FB task, while aTPJ displayed feedforward-like activity for the SA task. This pattern was further confirmed by a hierarchical cluster analysis. Overall, these findings suggest that the functional specialization, which is even reflected at the level of cortical layers, may enable TPJ to switch between detecting social cues externally and contemplating about them internally.

Keywords: social cognition, theory of mind, layer fMRI, top-down, bottom-up, false belief, social animations

Zusammenfassung

Theory of Mind (ToM) ermöglicht es uns über die mentalen Zustände und Überzeugungen anderer nachzudenken und ist somit eine wichtige soziale Fähigkeit des Menschen. Neuroimaging-Studien zeigen, dass die *Temporo-Parietal Junction* (TPJ) sowie ein ganzes Netzwerk weiterer Gehirnregionen während des Nachdenkens über die Überzeugungen anderer besonders aktiviert sind. Zwei gängige Aufgaben, die zur Untersuchung von ToM verwendet werden, sind die *False Belief* (FB)-Aufgabe und die *Social Animations* (SA)-Aufgabe, von denen angenommen wird, dass sie je vor allem einen posterioren Teil (pTPJ) und einen anterioren Teil (aTPJ) der TPJ aktivieren. Während bereits mögliche Erklärungen für diese funktionelle Spezialisierung vorgeschlagen wurden, bleiben die genauen Wirkmechanismen noch größtenteils ungeklärt. In der vorliegenden Studie wurde daher die neuronale Aktivität in den kortikalen Schichten von pTPJ und aTPJ während der FB- und SA-Aufgabe mit hochauflösendem Laminar-fMRT untersucht. Erstens bestätigen die Ergebnisse die Wichtigkeit des ToM-Netzwerks und insbesondere der TPJ für das Nachdenken über die mentalen Zustände und Überzeugungen anderer. Zweitens wurde eine signifikante Interaktion zwischen den Aufgaben und Regionen festgestellt, welche die erwartete funktionelle Spezialisierung der einzelnen Regionen für die verschiedenen Aufgaben unterstreicht. Drittens wiesen die Profile der kortikalen Schichten in beiden Aufgaben auf eine „Feedback-artige" Aktivität hin. Aufgeteilt nach Region jedoch, zeigte pTPJ eine „Feedback-artige" Aktivität für die FB-Aufgabe und aTPJ eine „Feedforward-artige" Aktivität für die SA-Aufgabe. Dieses Muster wurde auch durch eine hierarchische Clusteranalyse bestätigt. Insgesamt deuten diese Ergebnisse darauf hin, dass die funktionelle Spezialisierung, die sich sogar in den kortikalen Schichten widerspiegelt, es der TPJ ermöglichen könnte, zwischen dem Erkennen externer sozialer Hinweise und dem internen Nachdenken über diese Hinweise zu wechseln.

Schlüsselwörter: Soziale Kognition, Theory of Mind, Laminar-fMRT, Top-Down, Bottom-Up, False Belief, Social Animations

Table of Contents

Introduction

Social Cognition

As famously recognized by Albert Einstein, “*When we survey our lives and endeavors, we soon observe that almost the whole of our actions and desires are bound up with the existence of other human beings. We see that our whole nature resembles that of social animals.*” (Einstein, 1954). Indisputably, humans organize their lives in inherently interweaved social structures (Nowak, 2006; Pinker, 2010; Richerson & Boyd, 1998). Interactions with family, friends, collaborators, or competitors are characterized by a variety of sophisticated social behaviors such as communication, cooperation, deception, encouragement, and intimidation. These interactions all require an adequate representation of past, interpretation of present, and a prediction of future behaviors of others. Social cognition, the ability to build representations of one's relationships with others and to use those abstract concepts to guide social behavior in a flexible way (Adolphs, 2001), is assumed to account for a large percentage of human cognition and hence hypothesized as a major driving force for both the brain’s phylogenetic and epigenetic development (Dunbar, 1998; Dunbar & Shultz, 2007). As a result, the *social predictive brain* (Brown & Brüne, 2012; Koster-Hale & Saxe, 2013) is constantly challenged to identify, manipulate, and integrate a vast spectrum of socially relevant cues from various sources of information. The origin of this information is traditionally assumed to be *top-down,* representation-driven (e.g., knowing something) or *bottom-up,* stimulus-driven (e.g., perceiving something).

Theory of Mind

As humans live in large social ensembles, the brain needs to form representations not only within the own mind but also across individuals concerning the minds of others. Thus, complex social behaviors often require the capability to attribute and reason about mental states of others that are not directly observable (Frith & Frith, 2006, 2007; Frith & Frith, 2003) and especially to distinguish them from one’s own (van Veluw & Chance, 2014). This ability is termed *Theory of Mind* (ToM). Research on ToM was ignited by the pioneering work of Premack and Woodruff (1978) on chimpanzees’ implicit assumptions about the behavior of others and later transferred to human cognition through the contributions of Dennett (1978) and Wimmer and Perner (1983). Social cognition comprises ToM – while the former often involves observable, socially relevant stimuli (e.g., actions, facial expressions, and gaze direction), the latter is specifically concerned with unobservable mental states (Carrington & Bailey, 2009). Evidence from developmental psychology suggests that substantial ToM abilities emerge in childhood during preschool age (Wellman et al., 2001;

Wimmer & Perner, 1983) and continue to fluctuate across the lifespan (Blakemore, 2008). First evidence for an arguably implicit form of ToM can be already found in infants (Baillargeon et al., 2016). According to current evidence, children start to use mental state words at ~3 years old (Frith & Frith, 2003) and show an explicit understanding of another agent's false belief at ~4 years old (Baron-Cohen et al., 1985; Gopnik & Astington, 1988; Perner et al., 1987; Wellman et al., 2001; Wimmer & Perner, 1983). Moreover, numerous neurodevelopmental and psychiatric disorders are associated with disrupted development of ToM abilities (Korkmaz, 2011). Especially individuals diagnosed with *Autism Spectrum Disorder* (ASD) were found to experience difficulties in understanding the mental states of others (Baron-Cohen et al., 1985).

Intriguingly, the literature on ToM is traversed by a strong theoretical dualism. Firstly, frameworks on abstract theorizing about mental states (Gopnik & Wellman, 1994) rival accounts of concrete biological simulation processes (Gallese & Goldman, 1998; Keysers & Gazzola, 2007). Secondly, controversies around implicit and explicit forms have emerged (Apperly & Butterfill, 2009; Heyes & Frith, 2014; Van Overwalle & Vandekerckhove, 2013). Implicit ToM is assumed to be a lower-level and thereby automatic, rapid, and language-free process that can be observed already in very young children and specific nonhuman species. Explicit ToM is assumed to be a higher-level and thereby effortful, slower, and language-based process that can be found in typically developing older children and adults. Here, the almost simultaneous development of language (Kobayashi et al., 2007) and self-control (Adolphs, 2001) are hypothesized to guide the emergence of belief reasoning in childhood. Thirdly, cognitive and affective notions of social cognition and ToM have been discussed (Abu-Akel & Shamay-Tsoory, 2011; Poletti et al., 2012), which may further reflect the broad content of possible social behaviors.

Brain Regions and Processes

Generally, the umbrella term ToM includes various socio-cognitive abilities involving perception and attention (e.g., perceiving socially relevant cues), memory and language (e.g., retrieving social concepts), executive functions (e.g., distinguishing and tracking intentions), as well as emotion processing (e.g., empathy; Korkmaz, 2011). In a divide-and-conquer strategy, cognitive neuroscience aims to map these functional building blocks to anatomical brain regions and thereby tackle and explain this overarching, abstract, and psychological construct step-by-step (Schaafsma et al., 2015). A growing body of literature postulates a specialized ToM network (see Figure 1) that consists of brain regions indicating specific and reliable neuronal activation patterns for a range of tasks and stimuli (Schurz & Perner, 2015). While exact definitions and criteria of this network vary, regions such as *Temporo-Parietal Junction* (TPJ), *medial Prefrontal Cortex* (mPFC), and the *Temporal Poles* (TP) have been commonly identified as key network nodes (Frith & Frith, 2006; Frith & Frith, 2003; Kliemann et al., 2008; Mitchell, 2007; Saxe & Powell, 2006; Saxe & Wexler, 2005; Scholz et al., 2009; Young et al., 2010; Young et al., 2007). However, it is important to note that ToM may emerge solely through the coordinated synergy of these regions and subprocesses (Schaafsma et al., 2015). While the majority of the literature agrees on the involvement of these specific brain regions in social cognition, their underlying computational processing mechanisms remain highly debated (Baker, 2012; Frith, 2012; Gallagher & Frith, 2003; Kilner et al., 2007; Koster-Hale & Saxe, 2013; Yoshida et al., 2008). After all, the ToM network may be considered merely as a simplified but influential working model. The following overview aims to briefly introduce the core network regions together with their proposed functionality, focusing especially on TPJ as the subject of investigation of this thesis.

Temporo-Parietal Junction

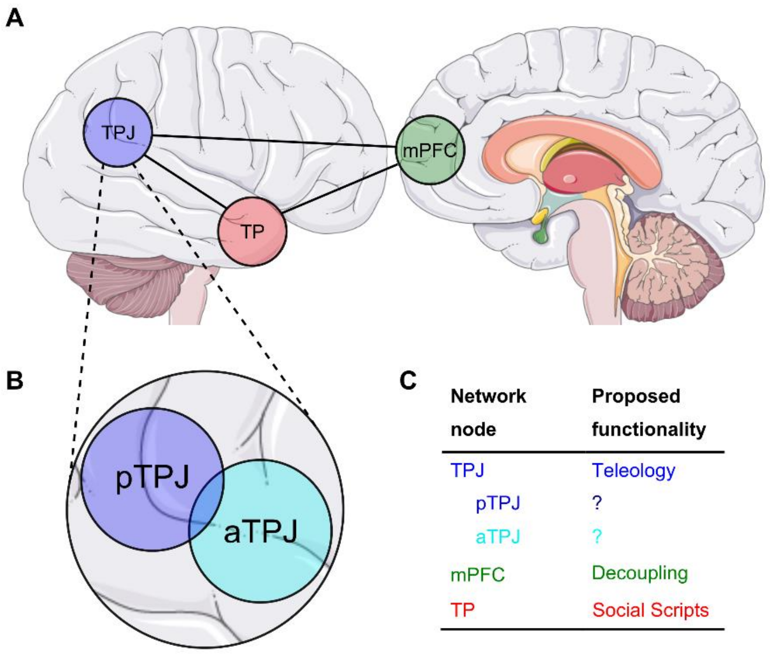
TPJ is located roughly at *Inferior Parietal Lobe* (IPL; BA39+40) and marks the transition of the temporal and parietal cortices. It is often divided into an anterior part (aTPJ) and a posterior part (pTPJ; Schurz et al., 2014). In the ToM literature, TPJ is commonly reported to respond selectively to mental state reasoning and in turn often highlighted as the most important brain region for this ability (Kobayashi et al., 2007; Sommer et al., 2007). It is associated with perspective taking in means-end reasoning about the intentions of others' actions, also termed "*teleology*" (Apperly et al., 2004; Frith & Frith, 1999; Perner & Leekam, 2008). Further evidence stresses its role in the detection of agency (Frith & Frith, 2003) and biological motion (Allison et al., 2000; Puce & Perrett, 2003; Saygin, 2007) as well as the extraction of social information from external stimuli (Carter et al., 2012; Mars et al., 2011; Scholz et al., 2009). Interestingly, this might be in line with other literature identifying especially the right TPJ as a hub for attention reorienting and switching between externally and internally oriented attention networks (Bzdok et al., 2013; Corbetta et al., 2008; Decety & Lamm, 2007; Mitchell, 2007; Scholz et al., 2009). To understand the role of TPJ, it is worth noting that mental state reasoning differs from simple perspective taking as it involves a fundamental understanding that knowledge is dependent on experience (Wimmer et al., 1988). As the knowledge of other agents can be cognitive, connotative, or affective, ToM processes can revolve around beliefs, desires, or emotions alike (Frith & Frith, 2006). Moreover, as mental states are usually not overtly observable, they must be covertly inferred. As teleology is necessary for all forms of ToM, it can be probed throughout various experimental paradigms (Schurz et al., 2014).

Medial Prefrontal Cortex and the Temporal Poles

The ToM network further includes mPFC (BA10) and the TP (BA38). Known for its role in executive control (e.g., planning, decision making, and dealing with conflict; Miller & Cohen, 2001), mPFC has been linked to decoupling mental states from reality in numerous ToM studies (Amodio & Frith, 2006; Frith & Frith, 2003; Gallagher & Frith, 2003). Individuals with lesions in this region were found to be impaired in mental state reasoning and executive control (Apperly et al., 2004; Stuss et al., 2001; Young et al., 2010). Maintaining and accessing separate self- and other-generated representations of an unobservable mental state and an observable ground truth is essential especially in the case of a false belief. As the *medial Temporal Lobe* and the TP are known to be strongly involved in primarily declarative long-term memory (Herlin et al., 2021; Squire & Zola-Morgan, 1991), they are assumed to retrieve knowledge about previously acquired social scripts (Frith & Frith, 2003). Social scripts are the protocols that define unique social interactions. The well-known "*restaurant script*" (Schank & Abelson, 1977), for example, is composed of guests arriving at a restaurant, choosing a table, checking the menu, ordering food, etc. It demonstrates that social scripts are often closely related to declarative semantic or episodic (i.e., language-dependent) knowledge but can likewise consist of procedural (i.e., language-free) knowledge. Fascinatingly, the TP seem to be the only brain region that selectively responds to knowledge about complex social behaviors (Zahn et al., 2007) and are often found to be engaged in forms of ToM that require comparing conceptual knowledge to sensory input.

Figure 1

ToM Network



Network node	Proposed functionality
TPJ	Teleology
pTPJ	?
aTPJ	?
mPFC	Decoupling
TP	Social Scripts

Note. **(A)** Commonly defined ToM network nodes: TPJ, mPFC, and the TP. **(B)** Division of TPJ into posterior (pTPJ) and anterior (aTPJ) clusters. **(C)** Proposed functionality of individual network regions. Parts of this figure were generated using Servier Medical Art licensed under a Creative Commons Attribution 3.0 unported license.

ToM Tasks

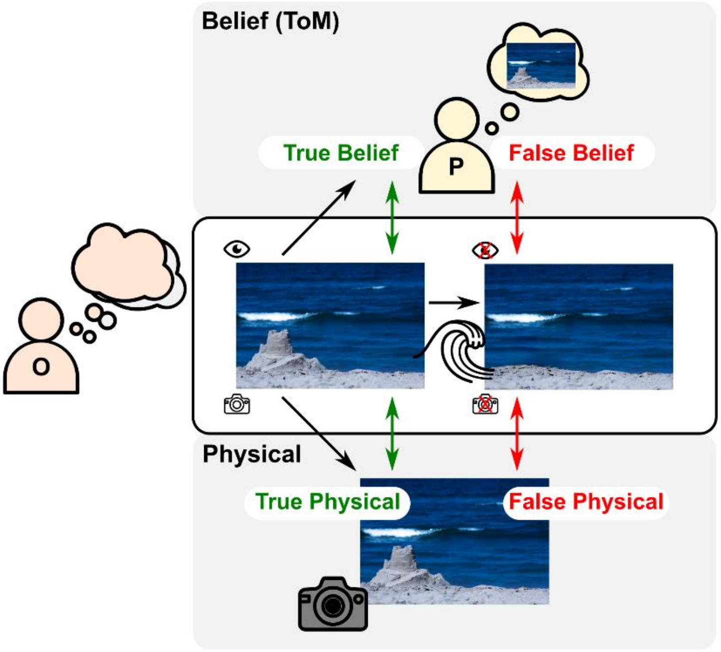
As ToM is a multifaceted construct, various experimental paradigms have been developed to probe its underlying subprocesses throughout different approaches. Commonly used tasks, among others, are *False Belief*, *Social Animations*, *Mind in the Eyes*, *Trait Judgements*, *Strategic Games*, and *Rational Actions* (Carrington & Bailey, 2009; Schurz et al., 2014; Schurz et al., 2021). The following section aims to provide a detailed description of the two particular paradigms that are compared in this thesis – *False Belief* and *Social Animations*.

False Belief Task

The flagship of ToM paradigms, the *False Belief* (hereafter FB) task, was proposed by Dennett (1978) and developed by Wimmer and Perner (1983) to investigate the emergence of belief reasoning in children. In the original version, children saw a protagonist place an object in location A and, after the protagonist left the scene, the object was moved to location B. Children, who witnessed the transfer, were then asked where the returning protagonist, who did not witness the transfer, would look for the desired object. This challenge requires perspective taking to predict the erroneous belief of the protagonist. Different versions of this task were used in subsequent years (e.g., Baron-Cohen et al., 1985; Saxe & Kanwisher, 2003) but especially the suggestion of a *False Photograph* story (Zaitchik, 1990), that involved changes in physically instead of mentally registered states, complemented the experiment with a closely matched control condition. Today, this “*Belief vs. Photograph*” (or more generally referred to as “*Belief vs. Physical*”) contrast (see Figure 2) is the most frequently used ToM paradigm in neuroimaging research (Wellman et al., 2001) and usually conveyed by logically structured written stories that are presented to participants inside the scanner (e.g., Dodell-Feder et al., 2011). In the experimental condition, a protagonist has a true belief that turns into a false belief as soon as the true physical state changes unknown to them. Then, the participants are usually asked to respond to a statement about the protagonist’s mental representation. The control condition has a similar structure, as a true physical state turns into a false physical state after being physically registered. Here, the only main difference lies in the outdated information, which is represented in a protagonist’s mind in the false belief and a physical registration in the false physical condition. In an item analysis of these stimuli, Dodell-Feder et al. (2011) showed that item-specific differences (e.g., lexical difficulty, logical complexity, and inhibitory demands) between the false belief and physical stories cannot explain activity differences between the conditions. Taken together, the FB task is a higher-level, language-based, theoretical probe, and based on a protagonist’s misinformation. In a meta-analysis of ToM tasks, Schurz et al. (2014) reported strongest activation for FB contrasts in bilateral TPJ, with peaks in pTPJ, as well as in mPFC, and *Precuneus*.

Figure 2

Theory of Mind and False Belief



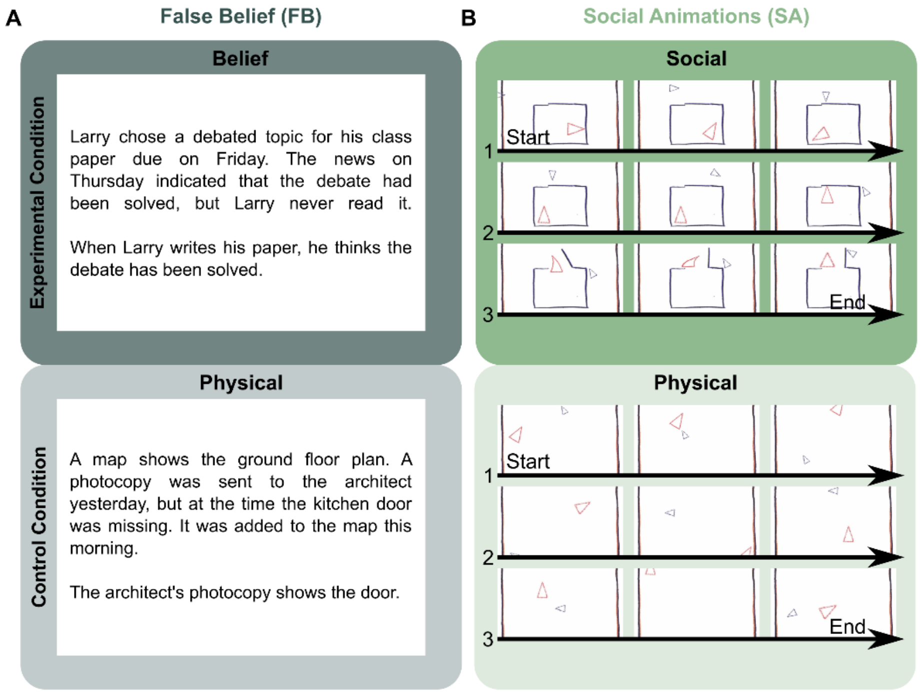
Note. Conceptual overview of ToM as an observer O's (usually a participant's) ability to infer and reason about a protagonist P's beliefs. While a true belief is congruent with the current state, a false belief is incongruent with the current state and often occurs following a change in the current state that happens known to O but unknown to P. The physical control condition follows the same logic with the only difference that, instead of P's belief, a physical representation (e.g., a photograph) registers the outdated information.

Social Animations Task

The *Social Animations* (hereafter SA) task was originally inspired by the work of Heider and Simmel (1944), who used simple two-dimensional movie displays of geometric shapes to investigate the emergence of higher-level percepts. Participants viewed the shapes move about and, regardless of instruction, reliably attributed internal states, personality traits, and emotions to the shapes. Instead of object identity, motion kinematics (e.g., temporal contingency and spatial proximity) were found to elicit these phenomenal relationships (Berry et al., 1992). Later studies confirmed that these subjective, higher-level percepts cannot be explained solely by their objective, lower-level retinal projections (Scholl & Tremoulet, 2000). The paradigm was revised by Castelli et al. (2000) with *Positron Emission Tomography* (PET) and yielded especially activations in the ToM network. In this setup, two triangles moved either corresponding to *complex intentional states* (e.g., surprising and mocking), *goal-directed* *actions* (e.g., dancing and chasing), or *random motion* (e.g., side-to-side and star-shaped). Relative changes of neural activity in ToM network regions were high for complex intentional states, intermediate for goal-directed actions, and low for random motion. This suggests that mental state reasoning may be possible through the interplay of automatic, perceptual attributions and effortful, conceptual interpretations. Following adaptations of the paradigm introduced novel conditions and stimulus variations (Martin & Weisberg, 2003). Taken together, the SA task is lower-level, largely perceptual, and automatic, while giving rise to higher-level cognitive concepts (e.g., sociality, agency, and causality; see Figure 3). As the SA task requires no misinformation, it may be argued that it probes more general social cognition instead of actual belief reasoning (Adolphs, 2001). Additionally, as this task is language-free, it is often used in neuroimaging studies investigating typically developing and ASD-diagnosed children (Abell et al., 2000; Ammons et al., 2018; Fitzpatrick et al., 2018; Kim et al., 2016; Klin & Jones, 2006; Vandewouw et al., 2021).

Figure 3

Example Stimuli



Larry chose a debated topic for his class paper due on Friday. The news on Thursday indicated that the debate had been solved, but Larry never read it.

When Larry writes his paper, he thinks the debate has been solved.

A map shows the ground floor plan. A photocopy was sent to the architect yesterday, but at the time the kitchen door was missing. It was added to the map this morning.

The architect's photocopy shows the door.

Note. **(A)** *False Belief* (FB) task example adapted from Dodell-Feder et al. (2011) with "*Belief vs. Physical*" contrast. **(B)** *Social Animations* (SA) task example adapted from Castelli et al. (2000) with "*Social vs. Physical*" contrast.

Bipartite TPJ

As briefly mentioned earlier, TPJ is often divided into a posterior (pTPJ) and an anterior part (aTPJ; Bzdok et al., 2016; Bzdok et al., 2013; Kernbach et al., 2018; Numssen et al., 2021). This dissociation is both functionally and anatomically motivated and further supported by two opposing views. According to one view, TPJ is a domain-specific ToM module (Saxe & Powell, 2006). Along these lines, pTPJ and aTPJ are subdivisions that may engage in different, rather independent variations of ToM tasks. However, according to another view, differences between the clusters are the result of a gradient in domain-general mechanisms (Cabeza et al., 2012). Thus, common computations in pTPJ and aTPJ underlie higher-level (i.e., ToM) and lower-level cognition (i.e., attention reorienting) alike (Decety & Lamm, 2007). The following section aims to shed light on the distinct functional and anatomical characteristics of these clusters according to both accounts.

Functional Division

In studies using FB and SA tasks, different activation patterns were found for pTPJ and aTPJ clusters. As reported in a meta-analysis by Schurz et al. (2014), while pTPJ is especially activated in FB tasks, aTPJ is predominantly engaged in SA tasks. Moreover, in a similar task-constrained meta-analysis, Bzdok et al. (2013) demonstrated that pTPJ may process predominantly internal information and communicate with a parietal network for social cognition and memory retrieval, and aTPJ may process especially external information and communicate with a midcingulate-motor-insula network for attention reorienting and saliency detection. Similarly, Corbetta et al. (2008) illustrated that the individual clusters may be embedded into separate attention networks. On the one hand, pTPJ may be part of a dorsal fronto-parietal network that closely resembles the *Default Mode Network* (DMN). The DMN is activated during internally directed complex thought such as retrieving autobiographical memory, envisioning the future, perspective taking, daydreaming, and mind wandering (Buckner et al., 2008; Raichle, 2015). Consequently, pTPJ may enable more abstract, endogenous, and representation-driven processing. Convincingly, a strong overlap between especially pTPJ, the ToM network, and the DMN has been reported (Mars et al., 2012; Mars et al., 2011; Schilbach et al., 2008). On the other hand, aTPJ may be part of a ventral fronto-parietal network and enable more concrete, exogenous, and stimulus-driven processing. This particular network is especially activated during attention reorienting. Thus, while pTPJ seems to be more involved in conceptual processes, aTPJ likely integrates perceptual processes to a greater extent. As mental state reasoning requires flexible shifts between externally directed perception (e.g., to capture observable social cues) and internally directed contemplation (e.g., to infer unobservable mental states), this functional specialization proposes an account of TPJ as a switching hub that links two antagonistic networks (Bzdok et al., 2013; Seghier, 2013). In accordance with these findings, Gobbini et al. (2007) reported selective pTPJ activation for reasoning about covert mental states and aTPJ activation for overt mental states. This fits well with studies that report engagement of aTPJ in biological motion detection (Allison et al., 2000; Puce & Perrett, 2003) and disrupted TPJ activity in lesions or *Transcranial Magnetic Stimulation* (TMS) leading to either out-of-body experiences (i.e., disrupted internal processing) or hallucinatory misperceptions (i.e., disrupted external processing) in respect to the disturbed region (Blanke & Arzy, 2005). In a meta-analysis, Decety and Lamm (2007) concluded that TPJ may solely switch between attention modes as it generates, tests, and corrects internal predictions of external events. They argue that higher-level cognitive abilities such as ToM may be entirely based on domain-general, low-level computations required to predict external events. However, as an interim summary, it should be noted that strong evidence exists to believe that TPJ may act as a gateway for linking internal and external information processing through functionally specialized clusters.

Anatomical Division

Generally, TPJ is known to receive input from the *Thalamus*, visual and auditory cortices, as well as the *Limbic System* and to output projections to mPFC and the TP among other regions (Decety & Lamm, 2007). While this circuitry clearly points to its embeddedness in the ToM network, it is still unclear to which extent, and in which order, the involved regions communicate to give rise to mental state reasoning. As insights into the cortical circuitry of TPJ can be obtained from cytoarchitectural data, Paquola et al. (2019) investigated cortical layer profiles with whole-brain *Magnetic Resonance Imaging* (MRI). Their results indicate that TPJ, which is part of heteromodal association cortex, receives and integrates input from both lower-level sensory as well as higher-level association areas. Intriguingly, the division of pTPJ and aTPJ also seems to hold in the case of various task-free, resting-state connectivity analyses (Bzdok et al., 2013; Caspers et al., 2011; Mars et al., 2011). This suggests that cytoarchitectural differences could give rise to the previously reported functional specialization. Therefore, while pTPJ may predominantly modulate information processing based on internally generated, top-down predictions in the case of covert false belief reasoning, aTPJ may be more incorporating external, bottom-up sensory cues in the case of overt social animations. Finally, evidence from functional and anatomical divisions immediately suggests that the functional specialization of the two clusters may be grounded in differently composed activations of cortical layers.

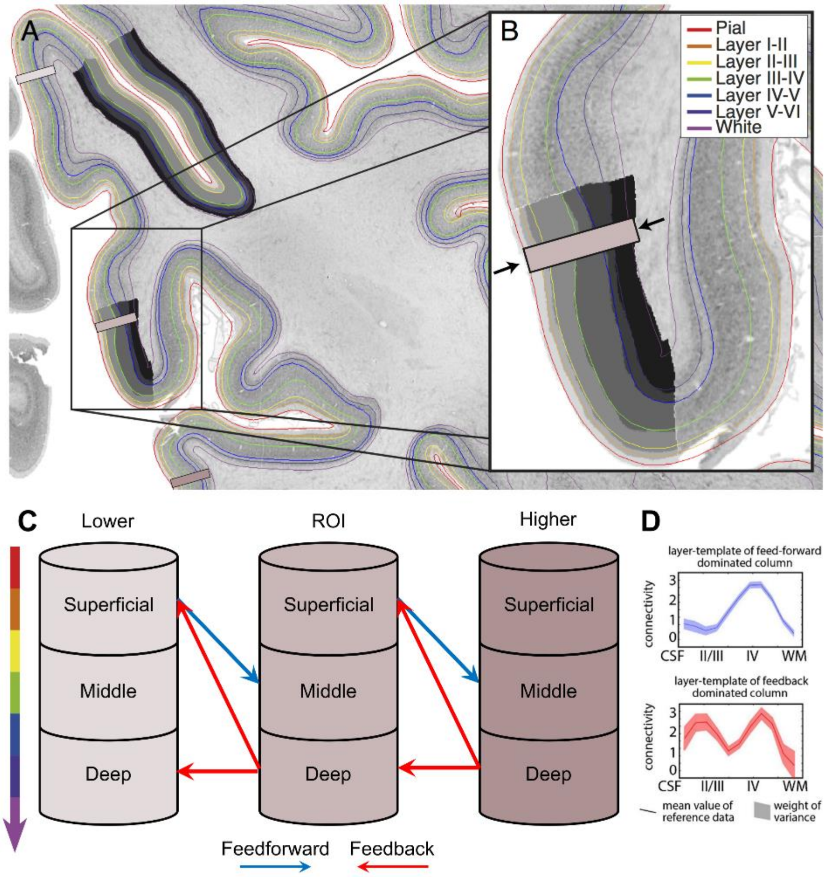
Cortical Organization

Cortical Circuits

Neocortex is hierarchically organized from macro- to microscale (Bastos et al., 2012; Van Essen & Maunsell, 1983). At macroscopic levels, it consists of cortical networks that are composed of individual regions linked by sparse but influential extrinsic projections through long-range white matter tracts. At mesoscopic levels, neurons with similar receptive field properties are vertically aligned in layers of canonical columns and interconnected through vast intrinsic connections (Douglas & Martin, 1991; Hubel & Wiesel, 1972). This short-range wiring makes up ~95 % of the cortical circuitry (Markov et al., 2011). According to the original cytoarchitectonic delineation by Brodmann (1909), neocortex can be divided into up to six cortical layers of ~0.2-1 mm in thickness. Gray matter (GM), which is composed of supragranular/superficial, granular/middle, and infragranular/deep cortical layers, is situated between cerebrospinal fluid (CSF) and white matter (WM) and characterized by its folding in sulci and gyri. Generally, these projections are grouped into the three following types (Felleman & Van Essen, 1991; Lamme & Roelfsema, 2000; Lamme et al., 1998; Rockland & Pandya, 1979). Firstly, feedforward projections connect lower- to higher-level areas, originate predominantly from superficial and terminate in middle layers. They often show specific receptive field properties and propagate strong driving excitatory activity (e.g., driving retinal input). Secondly, feedback projections connect higher- to lower-level areas, originate from deep and terminate outside of middle layers. They often modulate receptive field properties of feedforward projections through weaker but influential inhibitory activity (e.g., modulating cortical input). Thirdly, lateral projections connect areas within the same hierarchical level and enable recurrent processing unfolded and prolonged across time instead of space. Importantly, feedforward and feedback processes were found to activate different cortical layers (De Martino et al., 2015; Kok et al., 2016; Muckli et al., 2015; Scheeringa et al., 2016; Self et al., 2013). The prototypical feedforward profile is characterized by increased activity in middle layers representing a unimodal activation curve, while the prototypical feedback profile is characterized by increased activity in superficial and deep layers representing a bimodal activation curve (see Figure 4).

Figure 4

Cortical Layers



Note. **(A)** Part of neocortex adapted from the *BigBrain* 3D atlas of cortical layers (Wagstyl et al., 2020). **(B)** Gray matter (GM) consists of up to six layers that are often further grouped into superficial, middle, and deep bins. **(C)** Feedforward-dominated columns receive input from lower-level areas in middle layers and display a unimodal activation profile. Feedback-dominated columns receive input from higher-level areas in superficial and deep layers and display a bimodal activation profile. **(D)** Prototypical layer templates adapted from Huber et al. (2021b).

High-Resolution Layer fMRI

Over the last years, technological advances in *functional Magnetic Resonance Imaging* (fMRI) have increased the magnetic field strength of scanners and thereby pushed spatial resolution to submillimeter voxel sizes. This enables fMRI research to further advance into systems neuroscience. However, standard fMRI signals such as the *Blood Oxygen Level Dependent* (BOLD; e.g., see Logothetis & Wandell, 2004), which measures neural activity indirectly through blood oxygenation, are facing fundamental problems at this scale. Cerebral perfusion, and hence blood oxygenation, are distributed heterogeneously in both intra- and extravascular space due to large pial draining veins that bias the signal towards superficial layers and distort activity further away from the activated region (Duvernoy, 1999; Koopmans et al., 2011; Lawrence et al., 2019; Menon et al., 1995; Turner, 2002; Uludağ &

Blinder, 2018). Therefore, as layer fMRI aims towards more precise spatial resolutions, common signals show either decreased specificity across layers (e.g., GE- or SE-BOLD) or decreased sensitivity (e.g., T_2-prep or diffusion-weighted T_2-prep; Bandettini et al., 2021; Huber et al., 2018). An elegant solution to this problem is provided by the *Vascular Space Occupancy* (VASO) signal (Lu et al., 2003; Lu & van Zijl, 2012), which is a non-invasive, cerebral blood volume (CBV)-based measurement that shows the best sensitivity-specificity trade-off among common signals (Bandettini et al., 2021). VASO uses T_1-differences between blood and tissue to "null" and thereby strongly reduce vascular effects. More specifically, *Slice-Saturation Slab-Inversion VASO* (SS-SI-VASO; Huber et al., 2014) inverts the blood within a "slab", a specified cuboid region, through an inversion pulse and times the readout to the approximate recovery time of the longitudinal magnetization of blood (i.e., "blood-nulling time"). This way increased neural activity and hence CBV are reflected in a decreasing VASO signal (Huber et al., 2018; Lu & van Zijl, 2012). Moreover, as the negative VASO signal and other positive BOLD-specific components antagonize each other, the nulled and not-nulled signals are acquired in a near simultaneous but interleaved fashion to enable a post-hoc BOLD-correction of the nulled signal. While high-resolution layer fMRI is accompanied by challenges such as increasing measurement time, problematic motion artifacts, narrow coverage, decreasing signal-to-noise ratio, increasing thermal noise, required manual or semi-automatic layerification, and difficult between-subject analyses (Fedorenko, 2021; Finn et al., 2021; Huber et al., 2018; Merriam & Kay, 2022; Norris & Polimeni, 2019), recent advances have demonstrated that this method may even have a few compensating benefits in more common 3T settings (Huber et al., 2022).

Research Question and Hypotheses

This thesis aims to address the question whether differences in cortical feedback and feedforward processing in pTPJ and aTPJ can explain their functional specialization in FB and SA tasks through top-down and bottom-up mechanisms.

(1) *ToM Localizer*

Based on a fundamental body of literature, it is assumed that **(1)** TPJ is reliably activated in mental state reasoning.

(2) *Functional Specialization*

Moreover, it is expected that **(2a)** pTPJ indicates significantly increased BOLD signal changes in the FB task, as it is suggested to be predominantly engaged in covert mental state reasoning and internally directed information processing, and **(2b)** aTPJ indicates significantly increased BOLD signal changes in the SA task, as it is suggested to be predominantly engaged in overt mental state reasoning and externally directed information processing.

***(3)** Hierarchical Division of Cortical Layers*

Consequently, it is assumed that **(3a)** superficial and deep layers demonstrate significantly more pronounced activity in the FB task, reflecting predominantly feedback patterns as a proxy of top-down conceptual processes. Contrary, **(3b)** superficial and deep layers demonstrate significantly less pronounced activity in the SA task, as a proxy of more mixed top-down conceptual and bottom-up perceptual processes.

Material and Methods

Participants

The fMRI data were collected from a total of 20 healthy participants (12 female; 4 lefthanded) with an age ranging between 18 and 45 years (*M* = 25.6, *SD* = 6.12). All participants had normal or corrected-to-normal vision, were fluent German speakers, reported no history of psychiatric or neurological illnesses, and had no contraindications to MRI scanning. The experimental procedure was admitted by the local ethics committee, in accordance with the declaration of Helsinki, and agreed to by participants via written consent prior to the experiment. University students were compensated for taking part with accredited participation hours.

Stimuli and Experimental Design

The stimuli consisted of a total of 64 FB stories and 20 SA video clips. For the FB stories, 40 original stories were adapted (i.e., translated to German) from Dodell-Feder et al. (2011), who controlled for various linguistic properties, while 14 additional stories were generated in a similar fashion. The stories were presented in written form with white text on black background. Importantly, the stories were part of two conditions of equal number: **(1)** false belief stories (i.e., the experimental condition) and **(2)** false physical stories (i.e., the control condition), differing only in the medium of the falsely represented content. Moreover, each story was followed by a correct or an incorrect statement, to which participants were asked to respond to in a true-or-false choice. The number of correct and incorrect statements was balanced out. For the SA video clips, 15 original animations were adapted from Castelli et al. (2000) and 5 additional animations from Martin and Weisberg (2003). The video clips displayed simple geometric shapes move about on the screen and were individually adjusted to a duration of 15 s each. Similar to the FB stories, the animations were also part of two equally sized conditions: **(1)** 10 social animations (i.e., the experimental condition) and **(2)** 10 physical animations (i.e., the control condition), differing only in the sociality of movements. Participants were instructed to observe the video clips attentively.

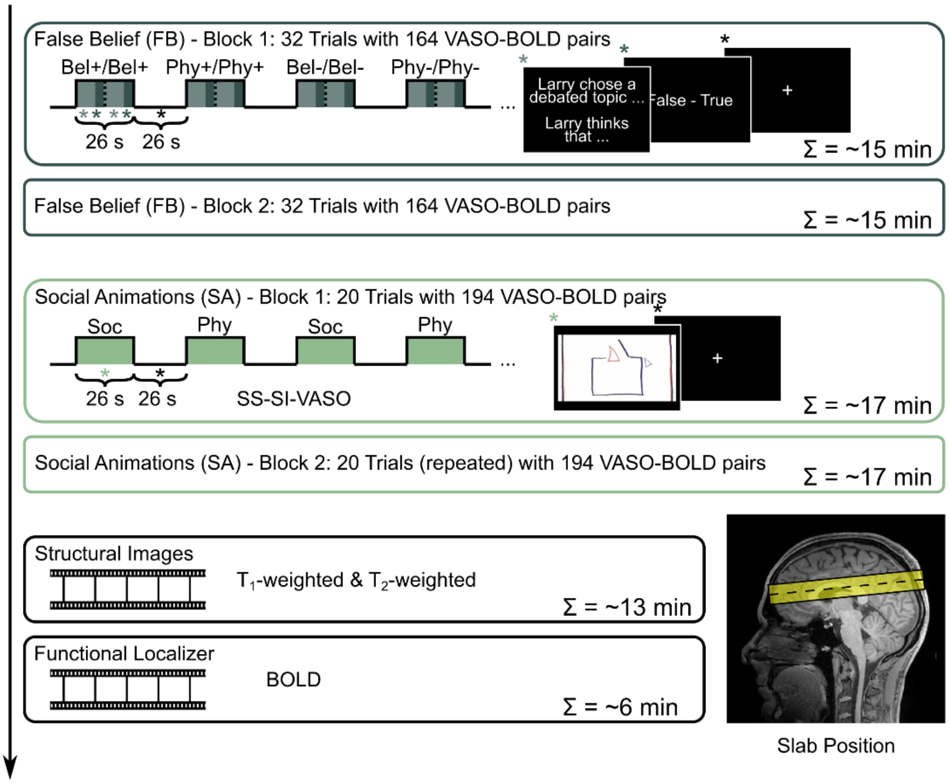
To begin with, inside the scanner, participants performed two FB story blocks. During each block, half of the stories were displayed. During each of these trials a fixation cross was presented for 26 s, followed by a pair of consecutive stories of the same condition (i.e., both false belief or both false physical) for 12 s each together with a correct/incorrect statement for 2 s each (see Figure 5). It took ~15 min to complete one of these blocks. Subsequently, participants watched two SA video clip blocks. During each block, all video clips were presented in random order but always alternated between the conditions (i.e., social and physical). During each of the trials, a fixation cross was presented for 26 s and followed by a video clip within a time window of 26 s. It took ~17 min to complete one of these blocks. In the next step, participants could either watch a movie or rest while structural T_1- and T_2-weighted images were obtained. This was completed after ~15 min. Finally, participants were instructed to watch a short animation movie (Sohn & Reher, 2009) commonly used as a functional localizer for ToM and empathy for pain (e.g., Jacoby et al., 2016). The movie had a duration of ~6 min.

fMRI Acquisition

Participants were scanned on a 3T Siemens MAGNETOM Prisma MRI scanner (Siemens AG, Erlangen, Germany) using a 64-channel RF transmit head coil of length 50 cm. Slab-selective, layer-specific images were acquired using a 3D-EPI sequence developed specifically for VASO imaging by Stirnberg and Stöcker (2021) with whole-brain MAGEC capabilities (Huber et al., 2021b). More precisely, the utilized SS-SI-VASO sequence by Huber et al. (2014) inverts the blood within a specified slab, which was precisely positioned above TPJ with an alignment in parallel to and between lateral sulcus and STS covering anterior and posterior clusters. The slab covered 30 slices and thus inflow of fresh blood into the slab was reduced during but possible in between inversion pulses. It was inverted uniformly with a global, adiabatic inversion pulse (e.g., Silver et al., 1984; Tannús & Garwood, 1997). The following parameters were applied for interleaved nulled and not-nulled imaging: FoV = 177 mm, FoV phase = 95.4 %, isotropic voxel size = 0.82 mm, TE = 27 ms, TR_{shot} = 64 ms, transmit reference voltage = 250 V, read bandwidth = 1092 Hz/Px, echo spacing = 1.02 ms, partial Fourier ky = 6/8, GRAPPA 3 (effective echo spacing = 0.47 ms, phase encoding bandwidth = 9.9 Hz/Px), TR_{vol} = 2075 ms, TR_{pair} = 5142 ms, inversion delay = 550 ms, flip angles = 33.1-60 °, water-selective excitation = binomial 1-1 pulses and bandwidth time product of 8. Whole-brain, structural T_1- and T_2-weighted images with 208 slices were obtained using the following parameters: FoV = 256 mm, FoV phase = 93,8 %, isotropic voxel size = 0.82 mm, TE_{T1} = 2.24 ms, TE_{T2} = 564 ms, TR_{T1} = 2400 ms, TR_{T2} = 3200 ms, transmit reference voltage = 250 V, read $bandwidth_{T1}$ = 210 Hz/Px, read $bandwidth_{T2}$ = 744 Hz/Px, echo $spacing_{T1}$ = 8.1 ms, echo $spacing_{T2}$ = 3.86 ms. Image reconstruction was performed using GRAPPA (Griswold et al., 2002).

Figure 5

Scanning Protocol


False Belief (FB) - Block 1: 32 Trials with 164 VASO-BOLD pairs
Bel+/Bel+
Phy+/Phy+
Bel-/Bel-
Phy-/Phy-
26 s
26 s
Larry chose a debated topic ...
Larry thinks that ...
False - True
+
Σ = ~15 min
False Belief (FB) - Block 2: 32 Trials with 164 VASO-BOLD pairs
Σ = ~15 min
Social Animations (SA) - Block 1: 20 Trials with 194 VASO-BOLD pairs
Soc
Phy
Soc
Phy
26 s
26 s
SS-SI-VASO
+
Σ = ~17 min
Social Animations (SA) - Block 2: 20 Trials (repeated) with 194 VASO-BOLD pairs
Σ = ~17 min
Structural Images
T1-weighted & T2-weighted
Σ = ~13 min
Functional Localizer
BOLD
Σ = ~6 min
Slab Position

Note. The scanning protocol consisted of two blocks of the FB task as well as two blocks of the SA task, in which interleaved nulled and not-nulled images were acquired, followed by structural images, and a functional localizer based on animated movie watching.

fMRI Preprocessing

The data were preprocessed using SPM12 (Functional Imaging Laboratory, University College London, UK), custom-made bash scripts (see Appendix), as well as commands from AFNI (Cox, 1996) and LayNii (Huber et al., 2021a).

BOLD – Functional Localizer

BOLD images acquired in the functional localizer task were motion corrected, normalized to the *Montreal Neurological Institute* (MNI) standard brain space, smoothed using a Gaussian filter with a full-width at half-maximum (FWHM) of 8 mm, and high-pass filtered to remove low-frequency noise. To model the data, a block design with boxcar regressors for the contrast onsets (e.g., "*Belief*" and "*Pain*") was convolved with a standard hemodynamic response function (HRF). These regressors were then compared in contrasts to estimate the model. No participants were excluded for this analysis.

VASO – FB and SA Task

To correct for motion, a mask was defined for each participant based on the combination of nulled and not-nulled images. This approach was used to reduce the influence of nonlinear distortions at high resolution and to only perform the transformations on the cortical tissue of interest. The nulled and not-nulled time series data were then realigned separately to this motion mask, as this approach has been shown to yield better results due to the different characteristics of the signals. To eliminate the influence of additional BOLD contamination on the nulled signal, the not-nulled signal was temporally upsampled by a factor of two. This was done to estimate the nulled signal at the time when the not-nulled signal was acquired to prevent any temporal differences in steady-state effects (Huber et al., 2014). Further dynamic division of the nulled by not-nulled signal yielded the desired, more specific SS-SI-VASO signal. To reduce noise amplification, this was performed on run-averaged time series data. Moreover, various quality checks based on statistical measures (e.g., mean, standard deviation, kurtosis, and skewness) were conducted.

Data Analysis

ROI Selection

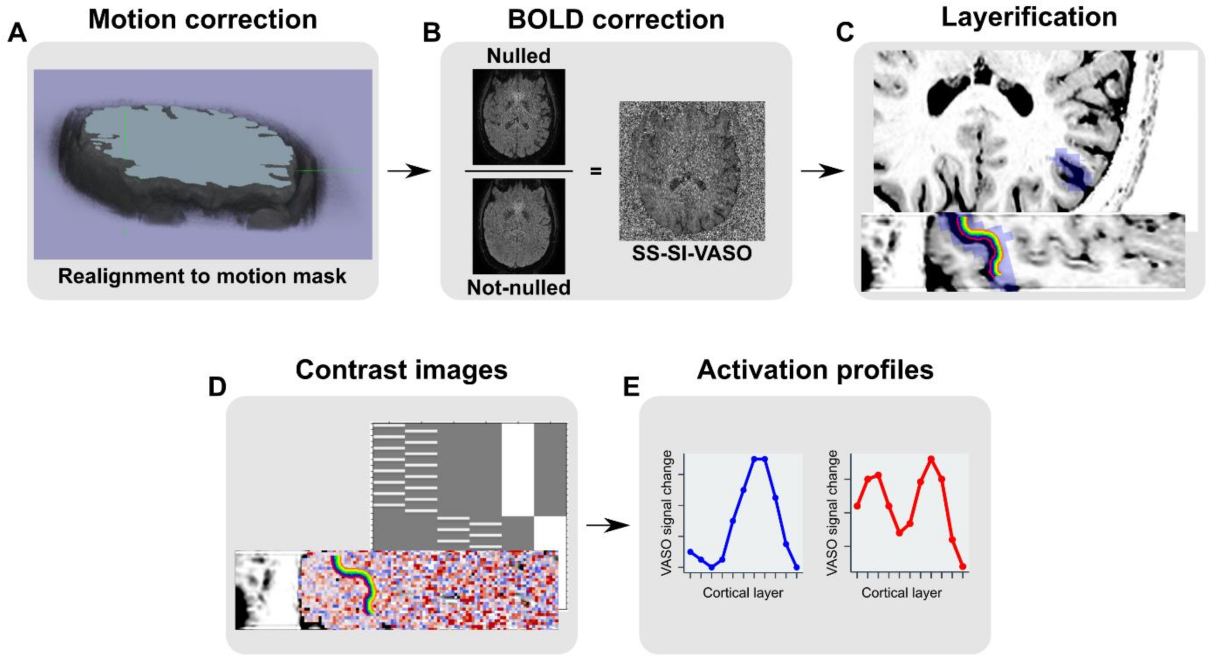
To select ROIs for the layer analyses, the results of the first-level analysis of the functional localizer were utilized. More specifically, the "*ToM > Pain*" contrast was applied, which had previously been demonstrated to isolate brain regions involved in belief and mental state reasoning (Jacoby et al., 2016). Significant clusters (p < .001) that were located near BA39 (*Angular Gyrus*) were used as an indicator of a participant's pTPJ. Preprocessed BOLD images and TPJ clusters of the functional localizer were then realigned to the BOLD and VASO data of the two main tasks. ROI rim files of the cortical patches were created manually in FSLeyes (Wellcome Centre for Integrative Neuroimaging, University of Oxford, UK). The position of pTPJ was approximated by the functional cluster in extension of the lateral sulcus on the posterior midline of the slab. The position of aTPJ was estimated to be more anterior and below the central midline. The resulting positions in unnormalized space were visually validated using characteristic anatomical landmarks based on the normalized center coordinates of pTPJ(x = 56, y = -56, z = 25) and aTPJ (x = 53, y = -31, z = 9) from a meta-analysis by Schurz et al. (2014). Notably, the top five conceptual associations for these cluster centers based on the meta-analytic *Neurosynth* tool (Yarkoni et al., 2011) are *theory of mind*, *mental states*, *beliefs*, *mind*, and *junction* for pTPJ and *auditory*, *superior temporal*, *speech*, *audiovisual*, and *temporal* for aTPJ. This ROI selection procedure was performed for both hemispheres and separately for the two tasks. To obtain a larger number of smoother layers and prevent resolution losses, the following steps were performed on spatially upsampled data (voxel size = 250 µm) instead of acquired resolution (voxel size = 0.8 mm). The manual delineation of WM and CSF boundaries was performed in sagittal slices on top of T_1-weighted border-enhanced images, which were created using a combination of nulled and not-nulled images. The selected cortical patches were chosen to have similar length, no holes, and an orientation that was ideally perpendicular to sagittal slices.

Layerification and Contrast Extraction

The ROI rim files were divided into layers using the automatic layering algorithm "LN_GROW_LAYERS" provided by LayNii. This process resulted in 11 layers with equal voxel depth, which represented ~9% of GM each in-between WM and CSF borders. Importantly, these layer bins should not be confused with cytoarchitectonic layers. Again, a block design with boxcar regressors for the contrast onsets of the two main tasks (i.e., "*Belief*" and "*Physical*" for FB, and "*Social*" and "*Physical*" for SA) was used to model the data. This design was convolved with a standard HRF before the regressors of the two tasks were compared using contrasts to estimate the model. In a first-level analysis, t-contrast images were computed for each task to represent the signal change of BOLD or VASO in each voxel. The contrast images of interest were then used to extract signal changes averaged over an entire ROI or its individual layers (see Figure 6) depending on the analysis. As the contrast images had already been generated at the single-participant level, no additional scaling of signal changes was applied. One participant was excluded from the analysis of BOLD signal changes averaged over entire ROIs due to poor signal quality. Additionally, one participant was excluded from the analysis of VASO signal changes within individual layers for the FB task, and two participants were excluded from this analysis for the SA task, due to excessive head motion (i.e., >1.6 mm) that would have led to highly inaccurate layer estimates. Within-subject comparisons were used for all statistical analyses of layer data, as they offer high sensitivity, functional resolution, and interpretability (Fedorenko, 2021). The extracted layer profiles were further compared to feedforward and feedback templates in a similar approach to Huber et al. (2021b) through template matching and between themselves in a hierarchical cluster analysis with a correlation-based dissimilarity measure (i.e., dissimilarity = 1 – *Pearson r*) based on scaled and centered data.

Figure 6

Analysis Pipeline



Note. Analysis pipeline to obtain the BOLD and VASO data of the main tasks after initial preprocessing steps. **(A)** Motion correction was performed on nulled and not-nulled images separately. **(B)** BOLD correction was performed through division of nulled by not-nulled images. **(C)** ROI rim files were manually drawn in a combined functional and anatomical approach. Layerification was performed using an automatic algorithm. **(D)** Contrast images reflecting signal change of BOLD or VASO were computed in first-level analyses for the main tasks. **(E)** Activation profiles were extracted for BOLD and VASO averaged over entire ROIs or individual layers.

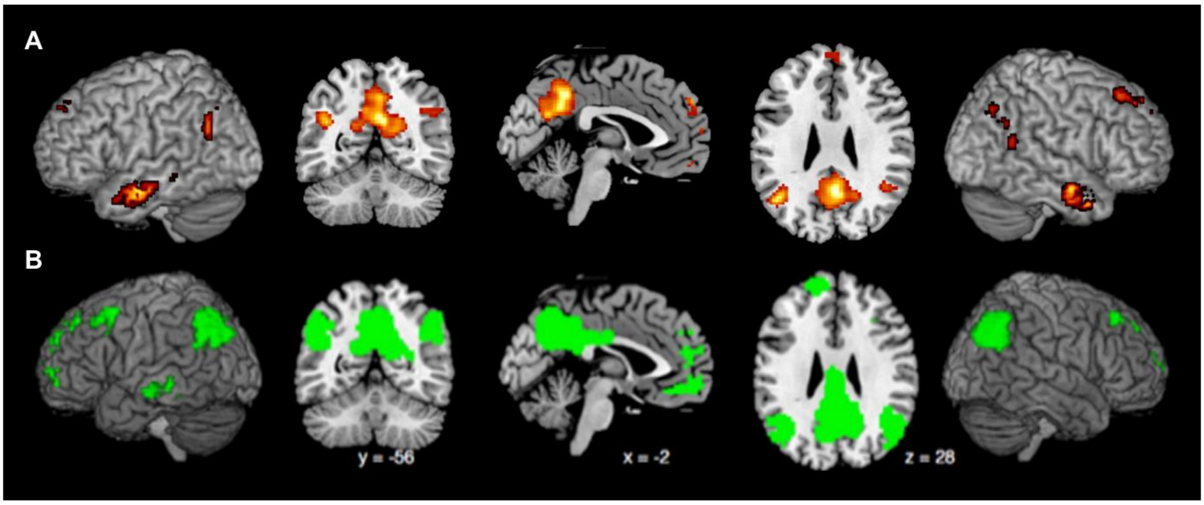
Results

Functional Localizer

First off, to localize TPJ and account for interindividual variability in functional specialization, a widely used functional localizer for ToM based on animated movie watching (Jacoby et al., 2016) was applied. As expected, significant clusters ($p_{FWE} < .05$) were found in all three ToM network regions, namely TPJ, mPFC, and an anterior part of *Medial Temporal Gyrus* (i.e., a region adjacent to the TP), for the “*Belief* > *Pain*” contrast. TPJ was indeed reliably more activated in reasoning about other’s mental states compared to their pain. Overall, the resulting clusters were strikingly similar to the original results by Jacoby et al. (2016; see Figure 7). However, in the present data, TPJ clusters were generally smaller in size and TP clusters were found bilaterally instead of only on the left hemisphere (see Table 1). These results corroborate the importance of TPJ in ToM and provide an important evidence base for the following hypotheses.

Figure 7

Group-Level Activation Pattern for the "Belief > Pain" Contrast of the Functional Localizer



Note. Whole-brain responses with significant clusters (p_{FWE} < .05, *k* = 0) based on the "*Belief* > *Pain*" contrast of the movie watching task. **(A)** The obtained results in the present experiment show activations bilaterally in *Precuneus*, *Angular Gyrus*, *Middle Temporal Gyrus*, and *Superior Medial Frontal Gyrus*. **(B)** Original results adopted from Jacoby et al. (2016) as a gold-standard comparison for the functional localizer.

Table 1

Significant Clusters for the "Belief > Pain" Contrast of the Functional Localizer

Cluster	Region	Nr. of voxels	x	y	z	Peak *t*
1	*Right Precuneus*	1222	3	-54	28	13.34
2	*Left Angular Gyrus*	156	-42	-60	25	12.38
3	*Left Middle Temporal Gyrus*	173	-60	-10	-15	12.08
4	*Right Middle Temporal Gyrus*	252	60	-7	-18	11.26
5	*Left Superior Medial Frontal Gyrus*	165	-2	56	35	9.76
6	*Right Angular Gyrus*	203	38	-64	40	9.69
7	*Left Middle Temporal Gyrus*	21	-60	-37	-2	7.80

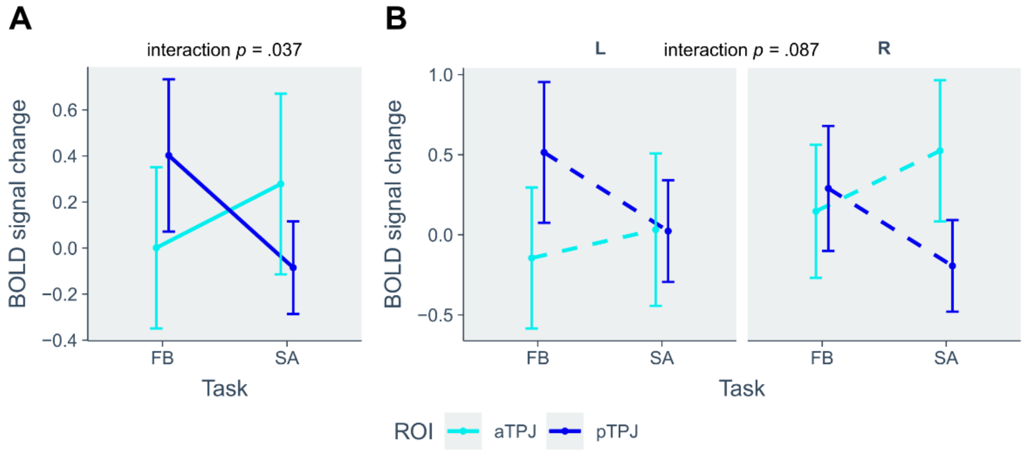
Note. Significant clusters (p_{FWE} < .05, *k* = 20). Labels of brain regions obtained from the *Neuromorphometrics* atlas. Peak coordinates of local clusters in MNI space.

Functional Specialization

Following up on this, to further investigate the functional specialization of pTPJ and aTPJ, signal changes averaged over entire ROIs were examined for the two main tasks. BOLD was used over VASO here, as no layer-specific resolution was required. Functional specialization was expected with pTPJ responding more to the main contrast of the FB task (i.e., “*Belief* > *Physical*”) and aTPJ responding more to the main contrast of the SA task (i.e., “*Social* > *Physical*”). To test this hypothesis, a two-way repeated measures ANOVA was conducted, as all necessary statistical assumptions were met. Without accounting for potential hemispheric differences between ROIs, no significant effects on overall BOLD signal changes were found for task ($F(1,17) = 0.42$; $p = .528$; *partial* $\eta^2 = .02$) and ROI ($F(1,17) = 0.05$; $p = .823$; *partial* $\eta^2 < .01$). This implied no overall activation differences between tasks and ROIs, respectively. However, as hypothesized, the interaction between task and ROI was found to have a significant, large effect ($F(1,17) = 5.11$; $p = .037$; *partial* $\eta^2 = .23$; see Figure 8A). Additional descriptive statistics supported the assumed directionality. On the one hand, pTPJ demonstrated increased, positive signal changes in the FB task ($M = 0.41$; $SD = 0.67$) compared to aTPJ ($M = 0.01$; $SD = 0.62$). On the other hand, aTPJ demonstrated increased, positive signal changes in the SA task ($M = 0.27$; $SD = 0.76$) compared to pTPJ ($M = -0.09$; $SD = 0.55$). Subsequent pairwise comparisons of ROIs, corrected for multiple comparisons (Benjamini & Hochberg, 1995), did not yield significant activation differences between ROIs in the individual tasks (FB: $t(17) = -1.79$; $p_{BH} = .091$ / SA: $t(18) = 1.80$; $p_{BH} = .089$). Taking into account potential hemispheric differences, no significant activation differences were found for task ($F(1,17) = 0.42$; $p = .528$; *partial* $\eta^2 = .02$), ROI ($F(3,51) = 1.67$; $p = .186$; *partial* $\eta^2 = .09$), or the interaction between the two ($F(3,51) = 5.04$; $p = .087$; *partial* $\eta^2 = .12$). The overall pattern of functional specialization remained consistent across hemispheres (see Figure 8B). Interestingly, strongest differences between pTPJ and aTPJ were observed in the left hemisphere for the FB task and the right hemisphere for the SA task.

Figure 8

Functional Specialization of TPJ Clusters



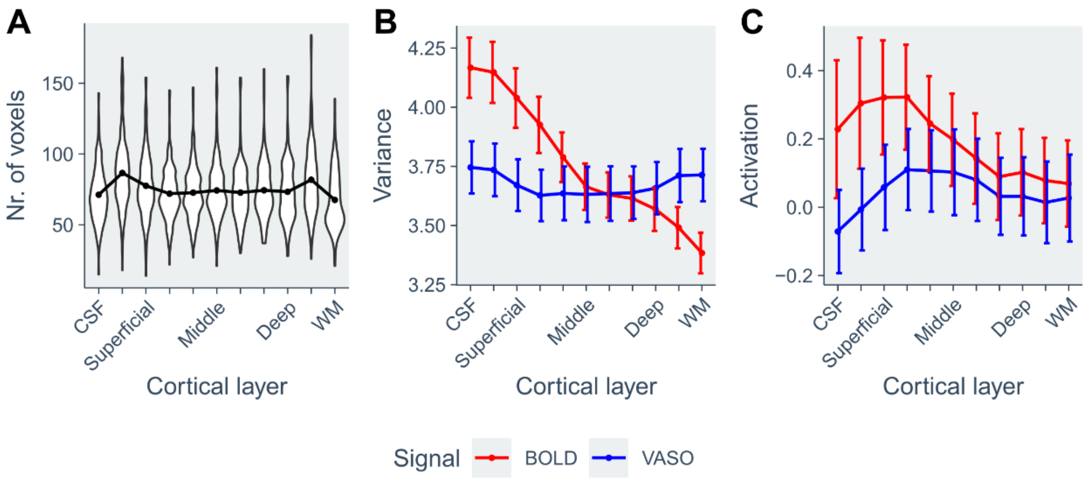
Note. BOLD signal changes for main task contrasts averaged over entire ROIs. Functional specialization **(A)** with and **(B)** without regard to hemispheric differences. Dots and error bars represent the mean and 95% confidence interval.

Cortical Circuits

Consequently, in order to assess whether VASO offers more specific results at the layer level compared to BOLD in the present data, key characteristics of the two signals were compared (see Figure 9). The number of voxels used for signal extraction varied notably between individual ROIs (i.e., across regions and participants) but remained relatively constant across individual layers with minor deviations observed in the second and penultimate layers due to curvature effects introduced by the sulci and gyri of GM. As anticipated, BOLD signal changes exhibited considerable variability in superficial layers due to heterogeneous blood contamination, while VASO signal changes were significantly more specific and reliable. On the other hand, in deeper layers, BOLD showed significantly lower variability compared to VASO. Overall, VASO displayed lower sensitivity compared to BOLD but unbiased results across layers with activation profiles similar to those in a comparable 3T layer fMRI study by Huber et al. (2022), making it suitable for the following layer analyses.

Figure 9

Characteristics of BOLD and VASO

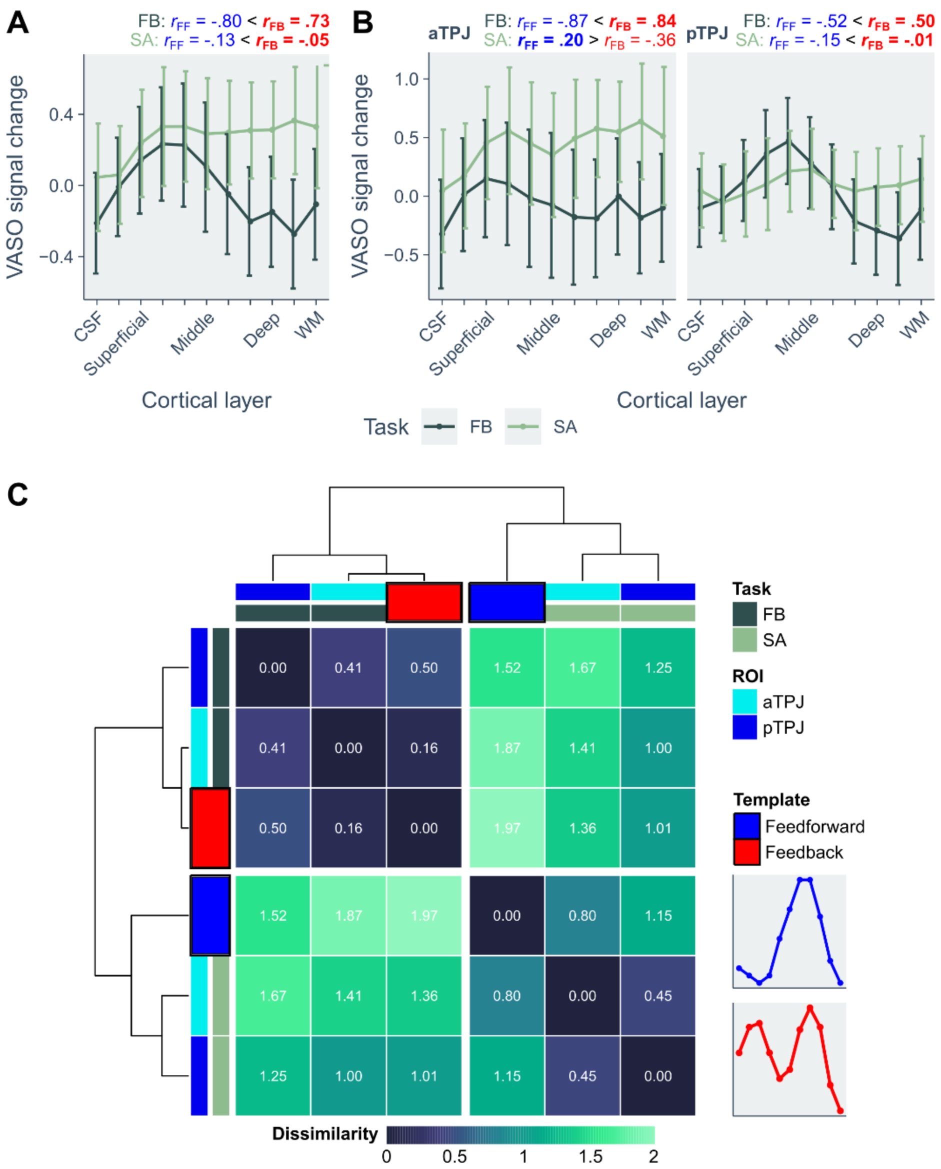


Note. Comparison of BOLD and VASO signals. **(A)** Number of voxels for individual layers included in ROIs. **(B)** Variance and **(C)** activation profiles across individual layers. Dots and error bars represent the mean and 95% confidence interval.

Finally, the unbiased VASO signal changes were used to compare the activation profiles of the two tasks. Based on the previous research on the various facets of ToM and the functionally specialized TPJ clusters involved, it was hypothesized that the FB task would primarily reflect feedback-like activity, while the SA task would show more mixed and unspecific activity. Interestingly, averaged over ROIs, layer profiles of the FB task showed increased activations in superficial and middle layers, while layer profiles of the SA task also indicated increased activity in deep layers (see Figure 10A). Comparisons with theoretical templates indicated that both tasks yielded an activation profile that was more similar to the feedback template as compared to the feedforward template. As expected, however, the FB task showed an activation profile that was much more similar to the feedback template (r_{FF} = -.80, r_{FB} = .73) with a pronounced peak in superficial and a suspected small peak in deep layers, compared to the SA task (r_{FF} = -.13, r_{FB} = -.05) with a rather mixed and unspecific profile. To gain further insight, layer profiles were separated by ROIs (see Figure 10B). The overall pattern remained similar and indicated predominantly feedback activity for the FB task in both pTPJ (r_{FF} = -.52, r_{FB} = .50) and aTPJ (r_{FF} = -.87, r_{FB} = .84) and slightly more influence of feedforward activity in the SA task for aTPJ (r_{FF} = .20, r_{FB} = -.36) but not pTPJ (r_{FF} = -.15, r_{FB} = -.01). This further strengthens the idea that pTPJ demonstrates a feedback-like profile in its specialized FB task, while aTPJ demonstrates a feedforward-like profile in its specialized SA task. Additionally, a hierarchical cluster analysis of layer profiles revealed two distinct clusters (see Figure 10C). Regardless of ROI, one cluster included the FB task and the feedback template, while the other cluster consisted of the SA task and the feedforward template. A complete cluster analysis, including hemispheric differences and additional task contrasts, can be found in the supplementary material (see Appendix – Supplementary Figure 2).

Figure 10

Layer Profiles



Note. **(A)** VASO layer profiles for main tasks averaged over ROIs and **(B)** divided by ROIs. Dots and error bars represent the mean and 95% confidence interval. The similarity of layer profiles to feedforward and feedback templates based on scaled and centered data is denoted by r_{FF} and r_{FB}, respectively. **(C)** Hierarchical cluster analysis of layer profiles, including feedforward and feedback templates adapted from Huber et al. (2021b).

Discussion

Conceptual Premises

As pointed out by Rauss and Pourtois (2013), "*one author's top-down may well be another one's bottom-up*". In this sense, for a fruitful and robust interpretation of the obtained results, it seems particularly important to adhere to a clearly defined terminology of processing mechanisms. On the one hand, the terms *top-down* and *bottom-up* are used to define the directionality of processing mechanisms on a larger scale, which can be either internally driven and based on a representational concept or externally driven and based on stimulus properties. On the other hand, these terms should not be confused with *feedback* and *feedforward*, which usually denote the anatomical circuitry underlying a neural sweep of activity on a much smaller scale. Consequently, abandoning this oversimplistic dichotomy, it is essential to note that solely the orchestrated interaction of feedforward, feedback, and recurrent connections may enable bottom-up and/or top-down processing after all (Rauschenberger, 2010). Therefore, the presented results require a careful interpretation and much more than just comparing the activation profiles to theoretical templates.

General Interpretation

To begin with, the results of the functional localizer underline the crucial role of the ToM network in mental state reasoning and reassure a correct localization of especially pTPJ. The observed group-level, whole-brain activation pattern indicated significant clusters overlapping with all three key ToM network regions, namely *Angular Gyrus* (i.e., TPJ), *Superior Medial Frontal Gyrus* (i.e., mPFC), and *Middle Temporal Gyrus* (i.e., a region adjacent to the TP), as well as *Precuneus*. Generally, this pattern fits well with numerous ToM neuroimaging studies (Frith & Frith, 2006; Frith & Frith, 2003; Kliemann et al., 2008; Mitchell, 2007; Saxe & Powell, 2006; Saxe & Wexler, 2005; Scholz et al., 2009; Young et al., 2010; Young et al., 2007). Moreover, it was also highly similar to that reported by Jacoby et al. (2016) in the original proposal of the functional localizer for ToM based on animated movie watching (*N* = 20, 12 females, all right-handed). However, in the present experiment, TPJ clusters were smaller in size, and temporal lobe clusters were observed more anterior as well as in both hemispheres instead of just the left hemisphere. While the processing of language is known to be generally left lateralized (Friederici, 2011, 2017), it is also linked to handedness (Ocklenburg et al., 2014). Therefore, the different results may be due to the inclusion of four left-handed participants in the current study, which could have contributed to the split of neural activation across hemispheres and in turn changed the overall pattern in the group-level analyses. As already discussed by Jacoby et al. (2016) and in line with the findings of several other authors (Bzdok et al., 2016; Bzdok et al., 2013; Corbetta et al., 2008; Mars et al., 2012;

Numssen et al., 2021; Schilbach et al., 2008; Yarkoni et al., 2011), the ToM network localized with the "*Belief > Pain*" contrast closely resembles the DMN, which is concerned with internal processing of complex thought, while the opposing "*Pain > Belief*" contrast highlights a saliency network (see Appendix – Supplementary Figure 1 and Supplementary Table 1), which is involved in external information processing. Nevertheless, in contrast to the ToM literature, the results did not demonstrate strong activations in mPFC, which could be due to the simplicity of the presented first-order beliefs in the animated movie that may not require a lot of decoupling (Amodio & Frith, 2006; Frith & Frith, 2003; Gallagher & Frith, 2003) and a contrast with pain scenarios instead of arguably more distinct physical states as a control condition (Dodell-Feder et al., 2011; Saxe & Kanwisher, 2003).

Based on these results, activations averaged over entire ROIs of pTPJ and aTPJ supported the notion that both are activated by mental state reasoning but, as hypothesized, the clusters demonstrate a functional specialization for the different tasks. Strikingly, while there were no overall activation differences between the individual tasks and ROIs, a significant interaction with a large effect size pointed towards a functionally bipartite TPJ. This observation is consistent with previous evidence on the specialization of pTPJ for FB tasks and aTPJ for SA tasks (Schurz et al., 2014). The fact that this effect was found even at the small size of the investigated cortical layer patches with only a few voxels demonstrates the assumed magnitude of functional specialization in TPJ. Moreover, differences between posterior and anterior clusters were smallest in the right hemisphere for the FB task and in the left hemisphere for the SA task. This implies additional language effects that could have enhanced differences between pTPJ and aTPJ of the same hemisphere in respect to the lateralization of language. This way, lateralization effects of language may increase the degree of functional specialization (i.e., activation differences between pTPJ and aTPJ) in the left, dominant hemisphere during the language-based FB task and in the right, non-dominant hemisphere during the language-free SA task. Promisingly, this observation is also in line with the previously reported overlap between ToM and language processing in left pTPJ (Bzdok et al., 2016; Mars et al., 2011; Seghier, 2013). This interpretation, however, will require further investigation, as it would also propose a stronger overall response to the FB task in the left pTPJ, which is usually observed in the right pTPJ more robustly (Perner et al., 2006; Saxe & Powell, 2006; Saxe & Wexler, 2005). Additionally, it seems important to also include the text-based physical control stories in this argumentation, since they are intended to control for such possible language effects. Taken together, the present results support the idea of a bipartite TPJ that could switch between internal and external information processing through flexible coordination of functionally specialized clusters (Bzdok et al., 2013; Corbetta et al., 2008; Gobbini et al., 2007; Seghier, 2013). Nevertheless, as it remains unclear on which specific mechanism this switching may be based on, the subsequent layer analyses aimed to elaborate on this matter.

The VASO signal clearly demonstrated its superiority over BOLD in the layer analyses, as it remained almost unbiased across the cortical depth. Notably, however, BOLD seemed to be more sensitive in the very deep layers indicating possible drawbacks of VASO's decreased sensitivity (Bandettini et al., 2021). Comparing layer profiles of both tasks averaged over ROIs indicated a main difference in deep layers where activations decreased considerably in the FB task and remained unchanged in the SA task. Averaged over ROIs, both tasks yielded a profile that was more similar to the feedback template. These results seem highly plausible in comparison with the histological whole-brain layer profiles reported by Paquola et al. (2019), as the entire posterior end of the lateral sulcus is characterized by feedback-like profiles and increases in profile ambiguity towards more anterior areas. Based on this argumentation, the layer profiles divided by ROIs and tasks clearly reflected the functional specialization, which seemed to be even recapitulated in the cortical circuitry. Here, pTPJ demonstrated feedback-like activity in the FB task, while aTPJ demonstrated feedforward-like activity in the SA task. The subsequent cluster analysis further supported this pattern by grouping the profiles based on their similarity into two distinct clusters that consisted of the two tasks and highlighted their characteristic directionality of processing mechanisms. Nevertheless, it should be noted that the observed profiles were overall very ambiguous and far from the theoretical templates. An example of this would be the missing peak in deep layers of pTPJ in both tasks, as reported for BA39 (i.e., the equivalent to pTPJ) by Finn et al. (2021). However, this observation could also be due to the previously noted low sensitivity of VASO in deep layers, which could have resulted in certain signal components not being accurately detected here. Additionally, the template correlations only give insights into whether a profile is more similar to either of the prototypical templates and hence cannot be interpreted in terms of overall strength (Huber et al., 2021b).

As these results hint towards the role of dissociable feedforward and feedback processes in the functional specialization of TPJ clusters, it could be argued that this cortical circuitry may serve as the basis for either internal vs. external processing (Bzdok et al., 2013), internal vs. external attention (Corbetta et al., 2008), prediction of external events (Decety & Lamm, 2007), or covert vs. overt mental state reasoning (Gobbini et al., 2007). Therefore, it could be promising to consider the ability of other populations or species to perform the two tasks. On the one side, ASD-diagnosed adults were found to have difficulties in the SA task, even though they seem to be less prominent compared to the FB task (Klin & Jones, 2006; Wilson, 2021), and ASD-diagnosed children show atypical activation patterns in several brain regions during the SA task, with the exception of TPJ (Vandewouw et al., 2021). Moreover, text- and cartoon-based FB tasks seem to activate the TPJ similarly in both adults and children (Kobayashi et al., 2007). Therefore, it cannot be argued directly that the SA task may require a markedly lower-level and implicit form of ToM, as TPJ may not be the most suitable region to investigate this. On the other side, while great apes and macaques were reported to possess precursors of ToM abilities (Hayashi et al., 2020; Kano et al., 2019), macaques do not seem to exhibit different gaze patterns in the social and physical conditions of the SA task (Schafroth et al., 2021). Similarly, the results by Roumazeilles et al. (2021) indicate also no different activation patterns in the macaque brain region that is equivalent to human TPJ in the social and physical conditions of the SA task. Interestingly, however, in comparison to human fMRI data, macaque TPJ shows more reliance on visuo-social information cues through stronger connectivity to lower-level medial STS and visual areas compared to human TPJ. Finally, recent evidence by Numssen et al. (2021), who compared attention reorienting, language processing, and belief reasoning in TPJ, may serve as the missing puzzle piece here. In accordance with the presented results, their functional parcellation of TPJ yielded two clusters with pTPJ being more active during belief reasoning and connected to the DMN, and aTPJ being less active during belief reasoning and connected to a ventral attention network. Their findings also suggest that the functional specialization of TPJ clusters exists in tasks related to attention reorienting, language processing, and belief reasoning alike, and that interactions of clusters in both hemispheres may depend on the specific task requirements. Taken together, it seems plausible that TPJ may serve as a hub for switching between abstract/internal and concrete/external information processing by flexibly accessing different brain networks that are reflected in the directionality of cortical processes at the layer level.

Alternative Interpretations

As the presented results may also fit other valid explanations, the following section aims to present and discuss a whole spectrum of alternative interpretations. It is not entirely clear to what extent FB and SA tasks capture the unique facets of ToM as proposed by the previous literature (Carrington & Bailey, 2009; Schurz et al., 2014; Schurz et al., 2021). Moreover, even if the functional specialization of pTPJ and aTPJ exists as expected, it remains challenging to draw conclusions from FB and SA tasks, as all mentioned candidate processes possibly interact and engage differently in both tasks. Eventually, it is possible that the "*Belief > Physical*" contrast in the FB task merely captures higher-level cognition instead of more top-down ToM processes and the "*Social > Physical*" contrast in SA tasks captures lower-level perception instead of more mixed ToM processes. In general, while this would certainly explain the observed pattern, it would not aid the investigation of neural processes underlying ToM.

Moreover, a fundamental body of literature underlines that social cognition and ToM are closely linked to language (Fodor, 1983; Lohmann & Tomasello, 2003; Milligan et al., 2007; Pyers & Senghas, 2009). As an example, language-based lateralization effects have been demonstrated for different forms of ToM tasks. Text-based and text-free tasks seem to more reliably activate left and right mPFC, respectively (Brunet et al., 2000; Fletcher et al., 1995; Gallagher et al., 2000; Goel et al., 1995). Transferring these findings to TPJ implies that mental state reasoning and language could share a similar functional specialization of conceptual and semantic properties. Additionally, other studies propose that the emergence of ToM may be directly tied to that of language (Lohmann & Tomasello, 2003). This way, more explicit and implicit forms of ToM are thought to require more and less language abilities, respectively (Apperly & Butterfill, 2009; Van Overwalle & Vandekerckhove, 2013). Some authors have even gone one step further to propose that "*explicit mind reading, like literacy, is a culturally inherited skill; it is passed from one generation to the next by verbal instruction.*" (Heyes & Frith, 2014). Either way, this reminds that the differences between the tasks could be largely influenced by, or even entirely based on, language effects. This means that language-independent differences between experimental and control conditions may exist separately in FB and SA contrasts, but still only language-dependent differences persist between the two tasks. Moreover, the contrast of a higher-level, language-based FB and a lower-level, language-free SA task may be overstated, as it is unclear how "nonverbal" the text-free SA task actually is. While it seems plausible that the tasks require language processes to a different extent, it may be problematic to treat the SA task as language-free. In contrast, it should be noted that there is also an established body of literature that suggests highly separated processing mechanisms for ToM and language in the brain (Amodio & Frith, 2006; Deen et al., 2015; Mar, 2011; Mason & Just, 2009; Paunov et al., 2019; Paunov et al., 2022; Shain et al., 2022; Varley & Siegal, 2000; Varley et al., 2001; Willems et al., 2011). Moreover, the specialization of the ToM network for mental state reasoning was even found to emerge gradually across development (Gweon et al., 2012; Richardson et al., 2018) and independent of language acquisition (Richardson et al., 2020).

Furthermore, evidence from visual neuroscience proposes a more detailed differentiation of feedback activity, as the peak in superficial layers was found to be associated with top-down attention (Lawrence et al., 2019), while the peak in deep layers was attributed to top-down prediction (Kok et al., 2016). This, however, would suggest especially strong differences in top-down predictions between the two tasks that do not fit the literature on ToM tasks (Schurz & Perner, 2015; Schurz et al., 2014).

Limitations

Given the ambitious endeavor to investigate a higher-level, socio-cognitive ability such as ToM with a novel methodological approach such as layer fMRI, it is plausible that several limitations may have confounded the obtained results. The first set of limitations involves the design of the study. Especially in the FB task, the rigid fMRI scanning protocol did not always enable ideal experimental conditions for the participants, as some of the presented stories were just too extensive for their limited presentation time. It may have been difficult for some participants to read all of the items carefully and get the gist of every story. This was due to the fact that too long trials would have resulted in mixing low frequency model effects and common low frequency noise. These challenging conditions may have resulted from the translation of the original stories by Dodell-Feder et al. (2011) from English to German, which could have led to much longer and more complex items. Besides that, as the name implies, FB stories only include false belief scenarios. However, mental state reasoning usually involves a variety of cognitive, connotative, and affective processes that are equally concerned with true beliefs or desires and emotions (Frith & Frith, 2006). This way, the presented evidence may only hold for a narrow spectrum of ToM abilities. Similarly, the long scanning time and repetitive presentation of SA videos, necessary for acquiring the layer fMRI data, may have contributed to some participants perceiving the task as tedious and exhausting, as inferred from some of their verbal reports. This of course could have affected the participants' attentiveness. Another imperfection of the SA task is that the original stimuli by Castelli et al. (2000) were created to fit three categories – *complex intentional states*, *goal-directed actions*, and *random motion*. In the present study the first two categories were grouped into a "*Social*" condition, which was further contrasted with the third category as a "*Physical*" condition. Compared to complex intentional states (e.g., triangle A surprises triangle B), goal-directed actions (e.g., triangles A and B are dancing) should not necessarily require mental state reasoning for an adequate interpretation. Taken together, these factors may have led to a general underestimation of ToM effects.

The second set of limitations involves the acquired measurements. While layer fMRI is usually performed at ultra-high magnetic field strength, recent advances by Huber et al. (2022) further justified its application in more widely accessible 3T settings. To date, however, no successful attempts to investigate ToM tasks in heteromodal association cortex have been made with layer fMRI. Interestingly, while VASO seems to be generally more preferable in 7T settings due to the increased signal-to-noise ratio, 3T settings possess their own advantages such as increased T_1-differences between blood and tissue, sharper images due to longer T_{2*} with less signal decay, and overall reduced artifact levels (Huber et al., 2022; Lu et al., 2013; Vu et al., 2017). Nevertheless, as argued by Merriam and Kay (2022), there are still at least three crucial challenges in the application of layer fMRI. The present study deliberately sacrifices signal-to-noise ratio to account for two of these challenges by applying VASO to control for common vascular biases (Duvernoy, 1999) and a 3T setting to reduce overall artifact levels. Its strongest limitation, however, may be best captured by the third challenge, which is an oversimplification of the cortical circuitry – especially in understudied regions beyond primary sensory cortex. According to Merriam and Kay (2022), these simplified models ignore important aspects of hierarchically distributed processing such as lateral and recurrent connections (Lamme & Roelfsema, 2000; Lamme et al., 1998), as well as other important aspects of neural dynamics (Arnal & Giraud, 2012), and therefore may lead to a rejection of unexpected results as "*messy*" data. Additionally, as the realignment of slab-restricted layer data to standard MNI coordinates is challenging, the ROIs had to be selected manually in unnormalized space. This may have introduced inaccuracies and/or biases that could have led to further spatial inaccuracies across ROIs and layers, as the selection of GM patches was rather difficult for some cortical patches due to imperfectly aligned GM (i.e., not always perpendicular to sagittal slices) and unclear CSF and WM boundaries. Future studies could circumvent these problems with whole-brain layer fMRI, which can be normalized to standard space, or the use of automatic delineation procedures (Huber et al., 2021a) and anatomical reference data (Numssen et al., 2021; Paquola et al., 2022; Paquola et al., 2019). This, in turn, would allow traditional ROI analyses (Poldrack, 2007) and aid the statistical power of the analyses by including more voxels and a more diverse spectrum of cortical tissue with different orientations. Furthermore, due to the use of within-subject designs in the statistical analyses applied to the data with high interindividual variability (Fedorenko, 2021), it may be challenging to generalize the findings to the overall population (Dubois & Adolphs, 2016). Here, between-subject analyses that compare different layers of individual participants (e.g., superficial layers from subject A and deep layers from subject B), to avoid dependencies within the layers of an ROI of a single participant, could offer a promising approach (Finn et al., 2021).

Open Questions and Future Directions

Future studies could investigate the origin of the observed feedforward inputs and feedback modulations in TPJ clusters with whole-brain layer fMRI or by incorporating methods such as seed-based connectivity analyses (e.g., Huber et al., 2021a). This could better reveal the orchestrated interplay of the different regions in the ToM network with TPJ as a pivotal point and further validate its role as a hub for switching between networks. Moreover, as feedforward and feedback connections propagate information predominantly in higher frequencies (e.g., gamma oscillations) and lower frequencies (e.g., alpha and beta oscillations) respectively (Bastos et al., 2012; Bosman et al., 2012; Buffalo et al., 2011; Maier et al., 2010), electrophysiological measures such as EEG and MEG could be used to further decipher the role of TPJ. Another important line of research may focus on the differences between the tasks. Similar to previous attempts to match FB stories and cartoons (Kobayashi et al., 2007), FB stories and SA videos could be extended through the use of closely matched scenarios of a wide range of social behaviors that can be presented in written stories and identically played out in videos of animated shapes. On top of that, other ToM tasks such as *Mind in the Eyes*, *Trait Judgements*, *Strategic Games*, and *Rational Actions* could be incorporated to detect possible task-related gradients in layer profiles of the TPJ clusters. Finally, comparisons with layer data from other populations such as ASD-diagnosed adults, patients with frontotemporal dementia, children, and nonhuman primates could provide further exciting insights.

Conclusion

The presented evidence suggests a functional specialization of TPJ, with pTPJ responding more to the FB task in feedback-like processes and aTPJ responding more to the SA task in feedforward-like processes. This interaction appears to increase in relation to language effects and implies that the requirements of the tasks may modulate the dynamic interplay of brain networks, rather than differences in the tasks or the ROIs alone. This way, pTPJ is activated in internal processing such as belief reasoning (e.g., Saxe & Kanwisher, 2003) to establish a connection between the ToM network and the DMN. Consequently, it may receive feedback from language and prefrontal areas such as mPFC during decoupling (Amodio & Frith, 2006). Meanwhile, aTPJ is involved in external processing such as the detection of agency (e.g., Frith & Frith, 2003) to establish a connection between the ToM network and a ventral attention network. As a result, it may receive feedforward activity from visual areas and feedback activity from temporal areas such as the TP during social script matching (e.g., Zahn et al., 2007) and the STS during biological motion perception (e.g., Saygin, 2007). This leads to the general assumption that the FB and SA task seem to have different requirements that make TPJ tap into different large-scale networks to switch between detecting social cues externally and contemplating about them internally (Bzdok et al., 2013; Corbetta et al., 2008; Gobbini et al., 2007; Numssen et al., 2021; Seghier, 2013). Some of these varying requirements of the FB and SA task could be related to the language-dependency of the material, the complexity and timescale of the scenarios, and additional cognitive mechanisms such as inhibitory control and memory. In conclusion, the presented results provide exciting new insights into the role of TPJ and may serve as a stimulating new foundation for future neuroimaging studies on ToM.

References

Abell, F., Happé, F., & Frith, U. (2000). Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development, 15*(1), 1-16. https://doi.org/10.1016/S0885-2014(00)00014-9

Abu-Akel, A., & Shamay-Tsoory, S. (2011). Neuroanatomical and neurochemical bases of theory of mind. *Neuropsychologia, 49*(11), 2971-2984. https://doi.org/10.1016/j.neuropsychologia.2011.07.012

Adolphs, R. (2001). The neurobiology of social cognition. *Current Opinion in Neurobiology, 11*(2), 231-239. https://doi.org/10.1016/S0959-4388(00)00202-6

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. https://doi.org/10.1016/s1364-6613(00)01501-1

Ammons, C. J., Doss, C. F., Bala, D., & Kana, R. K. (2018). Brain Responses Underlying Anthropomorphism, Agency, and Social Attribution in Autism Spectrum Disorder. *The open neuroimaging journal, 12*, 16-29. https://doi.org/10.2174/1874440001812010016

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7*(4), 268-277. https://doi.org/10.1038/nrn1884

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review, 116*(4), 953. https://doi.org/10.1037/a0016923

Apperly, I. A., Samson, D., Chiavarino, C., & Humphreys, G. W. (2004). Frontal and Temporo-Parietal Lobe Contributions to Theory of Mind: Neuropsychological Evidence from a False-Belief Task with Reduced Language and Executive Demands. *Journal of Cognitive Neuroscience, 16*(10), 1773-1784. https://doi.org/10.1162/0898929042947928

Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences, 16*(7), 390-398. https://doi.org/10.1016/j.tics.2012.05.003

Baillargeon, R., Scott, R. M., & Bian, L. (2016). Psychological Reasoning in Infancy. *Annual Review of Psychology, 67*, 159-186. https://doi.org/10.1146/annurev-psych-010213-115033

Baker, C. L. (2012). *Bayesian theory of mind: Modeling human reasoning about beliefs, desires, goals, and social relations*. [Doctoral Thesis, Massachusetts Institute of Technology].

Bandettini, P. A., Huber, L., & Finn, E. S. (2021). Challenges and opportunities of mesoscopic brain mapping with fMRI. *Current Opinion in Behavioral Sciences, 40*, 189-200. https://doi.org/10.1016/j.cobeha.2021.06.002

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

Bastos, Andre M., Usrey, W. M., Adams, Rick A., Mangun, George R., Fries, P., & Friston, Karl J. (2012). Canonical Microcircuits for Predictive Coding. *Neuron, 76*(4), 695-711. https://doi.org/10.1016/j.neuron.2012.10.038

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*(1), 289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Berry, D. S., Misovich, S. J., Kean, K. J., & Baron, R. M. (1992). Effects of Disruption of Structure and Motion on Perceptions of Social Causality. *Personality and Social Psychology Bulletin, 18*(2), 237-244. https://doi.org/10.1177/0146167292182016

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

Blanke, O., & Arzy, S. (2005). The Out-of-Body Experience: Disturbed Self-Processing at the Temporo-Parietal Junction. *The Neuroscientist, 11*(1), 16-24. https://doi.org/10.1177/1073858404270885

Bosman, Conrado A., Schoffelen, J.-M., Brunet, N., Oostenveld, R., Bastos, Andre M., Womelsdorf, T., Rubehn, B., Stieglitz, T., De Weerd, P., & Fries, P. (2012). Attentional Stimulus Selection through Selective Synchronization between Monkey Visual Areas. *Neuron, 75*(5), 875-888. https://doi.org/10.1016/j.neuron.2012.06.037

Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Barth.

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

Brunet, E., Sarfati, Y., Hardy-Baylé, M.-C., & Decety, J. (2000). A PET Investigation of the Attribution of Intentions with a Nonverbal Task. *NeuroImage, 11*(2), 157-166. https://doi.org/10.1006/nimg.1999.0525

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences, 1124*, 1-38. https://doi.org/10.1196/annals.1440.011

Buffalo, E. A., Fries, P., Landman, R., Buschman Timothy, J., & Desimone, R. (2011). Laminar differences in gamma and alpha coherence in the ventral stream. *Proceedings of the National Academy of Sciences, 108*(27), 11262-11267. https://doi.org/10.1073/pnas.1011284108

Bzdok, D., Hartwigsen, G., Reid, A., Laird, A. R., Fox, P. T., & Eickhoff, S. B. (2016). Left inferior parietal lobe engagement in social cognition and language. *Neuroscience & Biobehavioral Reviews, 68*, 319-334. https://doi.org/10.1016/j.neubiorev.2016.02.024

Bzdok, D., Langner, R., Schilbach, L., Jakobs, O., Roski, C., Caspers, S., Laird, A. R., Fox, P. T., Zilles, K., & Eickhoff, S. B. (2013). Characterization of the temporo-parietal junction by combining data-driven parcellation, complementary connectivity analyses, and functional decoding. *NeuroImage, 81*, 381-392. https://doi.org/10.1016/j.neuroimage.2013.05.046

Carrington, S. J., & Bailey, A. J. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human Brain Mapping, 30*(8), 2313-2335. https://doi.org/10.1002/hbm.20671

Carter, R. M., Bowling Daniel, L., Reeck, C., & Huettel Scott, A. (2012). A Distinct Role of the Temporal-Parietal Junction in Predicting Socially Guided Decisions. *Science, 337*(6090), 109-111. https://doi.org/10.1126/science.1219681

Caspers, S., Eickhoff, S. B., Rick, T., von Kapri, A., Kuhlen, T., Huang, R., Shah, N. J., & Zilles, K. (2011). Probabilistic fibre tract analysis of cytoarchitectonically defined human inferior parietal lobule areas reveals similarities to macaques. *NeuroImage, 58*(2), 362-380. https://doi.org/10.1016/j.neuroimage.2011.06.027

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. *Social Neuroscience*, 155-169. https://doi.org/10.1006/nimg.2000.0612

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The Reorienting System of the Human Brain: From Environment to Theory of Mind. *Neuron, 58*(3), 306-324. https://doi.org/10.1016/j.neuron.2008.04.017

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, 29*(3), 162-173. https://doi.org/10.1006/cbmr.1996.0014

De Martino, F., Moerel, M., Ugurbil, K., Goebel, R., Yacoub, E., & Formisano, E. (2015). Frequency preference and attention effects across cortical depths in the human primary auditory cortex. *Proceedings of the National Academy of Sciences, 112*(52), 16036-16041. https://doi.org/10.1073/pnas.1507552112

Decety, J., & Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. *Neuroscientist, 13*(6), 580-593. https://doi.org/10.1177/1073858407304654

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. https://doi.org/10.1093/cercor/bhv111

Dennett, D. C. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences, 1*(4), 568-570. https://doi.org/10.1017/S0140525X00076664

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. https://doi.org/10.1016/j.neuroimage.2010.12.040

Douglas, R. J., & Martin, K. A. (1991). A functional microcircuit for cat visual cortex. *The Journal of Physiology, 440*, 735-769. https://doi.org/10.1113/jphysiol.1991.sp018733

Dubois, J., & Adolphs, R. (2016). Building a Science of Individual Differences from fMRI. *Trends in Cognitive Sciences, 20*(6), 425-443. https://doi.org/10.1016/j.tics.2016.03.014

Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews, 6*(5), 178-190. https://doi.org/10.1080/03014460902960289

Dunbar, R. I. M., & Shultz, S. (2007). Evolution in the Social Brain. *Science, 317*(5843), 1344-1347. https://doi.org/10.1126/science.1145463

Duvernoy, H. M. (1999). *The human brain: surface, three-dimensional sectional anatomy with MRI, and blood supply*. Springer Science & Business Media.

Einstein, A. (1954). *Ideas and Opinions* (C. Seelig, Ed.). University of Michigan.

Fedorenko, E. (2021). The early origins and the growing popularity of the individual-subject analytic approach in human neuroscience. *Current Opinion in Behavioral Sciences, 40*, 105-112. https://doi.org/10.1016/j.cobeha.2021.02.023

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1*(1), 1-47. https://doi.org/10.1093/cercor/1.1.1-a

Finn, E. S., Huber, L., & Bandettini, P. A. (2021). Higher and deeper: Bringing layer fMRI to association cortex. *Progress in Neurobiology, 207*, 101930. https://doi.org/10.1016/j.pneurobio.2020.101930

Fitzpatrick, P., Frazier, J. A., Cochran, D., Mitchell, T., Coleman, C., & Schmidt, R. C. (2018). Relationship Between Theory of Mind, Emotion Recognition, and Social Synchrony in Adolescents With and Without Autism. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.01337

Fletcher, P. C., Happé, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S. J., & Frith, C. D. (1995). Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. *Cognition, 57*(2), 109-128. https://doi.org/10.1016/0010-0277(95)00692-R

Fodor, J. A. (1983). *The modularity of mind*. MIT press.

Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiological Reviews, 91*(4), 1357-1392. https://doi.org/10.1152/physrev.00006.2011

Friederici, A. D. (2017). *Language in Our Brain: The Origins of a Uniquely Human Capacity*. The MIT Press. https://doi.org/10.7551/mitpress/11173.001.0001

Frith, C. D. (2012). The role of metacognition in human social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1599), 2213-2223. https://doi.org/10.1098/rstb.2012.0123

Frith, C. D., & Frith, U. (1999). Interacting Minds--A Biological Basis. *Science, 286*(5445), 1692-1695. https://doi.org/10.1126/science.286.5445.1692

Frith, C. D., & Frith, U. (2006). The Neural Basis of Mentalizing. *Neuron, 50*(4), 531-534. https://doi.org/10.1016/j.neuron.2006.05.001

Frith, C. D., & Frith, U. (2007). Social Cognition in Humans. *Current Biology, 17*(16), R724-R732. https://doi.org/10.1016/j.cub.2007.05.068

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*(1431), 459-473. https://doi.org/doi:10.1098/rstb.2002.1218

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences, 7*(2), 77-83. https://doi.org/10.1016/S1364-6613(02)00025-6

Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia, 38*(1), 11-21. https://doi.org/10.1016/S0028-3932(99)00053-6

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences, 2*(12), 493-501. https://doi.org/10.1016/S1364-6613(98)01262-5

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. https://doi.org/10.1162/jocn.2007.19.11.1803

Goel, V., Grafman, J., Sadato, N., & Hallett, M. (1995). Modeling other minds. *NeuroReport-International Journal for Rapid Communications of Research in Neuroscience, 6*(13), 1741-1746. https://doi.org/10.1097/00001756-199509000-00009

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, 26-37. https://doi.org/10.2307/1130386

Gopnik, A., & Wellman, H. M. (1994). The theory theory. An earlier version of this chapter was presented at the Society for Research in Child Development Meeting, 1991.

Griswold, M. A., Jakob, P. M., Heidemann, R. M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., & Haase, A. (2002). Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic Resonance in Medicine, 47*(6), 1202-1210. https://doi.org/10.1002/mrm.10171

Gweon, H., Dodell-Feder, D., Bedny, M., & Saxe, R. (2012). Theory of Mind Performance in Children Correlates With Functional Specialization of a Brain Region for Thinking About Thoughts. *Child Development, 83*(6), 1853-1868. https://doi.org/10.1111/j.1467-8624.2012.01829.x

Hayashi, T., Akikawa, R., Kawasaki, K., Egawa, J., Minamimoto, T., Kobayashi, K., Kato, S., Hori, Y., Nagai, Y., Iijima, A., Someya, T., & Hasegawa, I. (2020). Macaques Exhibit Implicit Gaze Bias Anticipating Others' False-Belief-Driven Actions via Medial Prefrontal Cortex. *Cell Reports, 30*(13), 4433-4444.e4435. https://doi.org/10.1016/j.celrep.2020.03.013

Heider, F., & Simmel, M. (1944). An Experimental Study of Apparent Behavior. *The American Journal of Psychology, 57*(2), 243-259. https://doi.org/10.2307/1416950

Herlin, B., Navarro, V., & Dupont, S. (2021). The temporal pole: From anatomy to function—A literature appraisal. *Journal of Chemical Neuroanatomy, 113*, 101925. https://doi.org/10.1016/j.jchemneu.2021.101925

Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading. *Science, 344*(6190), Article 1243091. https://doi.org/10.1126/science.1243091

Hubel, D. H., & Wiesel, T. N. (1972). Laminar and columnar distribution of geniculo-cortical fibers in the macaque monkey. *The Journal of Comparative Neurology, 146*(4), 421-450. https://doi.org/10.1002/cne.901460402

Huber, L., Finn, E. S., Chai, Y., Goebel, R., Stirnberg, R., Stöcker, T., Marrett, S., Uludag, K., Kim, S.-G., Han, S., Bandettini, P. A., & Poser, B. A. (2021b). Layer-dependent functional connectivity methods. *Progress in Neurobiology, 207*, 101835. https://doi.org/10.1016/j.pneurobio.2020.101835

Huber, L., Ivanov, D., Handwerker, D. A., Marrett, S., Guidi, M., Uludağ, K., Bandettini, P. A., & Poser, B. A. (2018). Techniques for blood volume fMRI with VASO: From low-resolution mapping towards sub-millimeter layer-dependent applications. *NeuroImage, 164*, 131-143. https://doi.org/10.1016/j.neuroimage.2016.11.039

Huber, L., Ivanov, D., Krieger, S. N., Streicher, M. N., Mildner, T., Poser, B. A., Möller, H. E., & Turner, R. (2014). Slab-selective, BOLD-corrected VASO at 7 Tesla provides measures of cerebral blood volume reactivity with high signal-to-noise ratio. *Magnetic Resonance in Medicine, 72*(1), 137-148. https://doi.org/10.1002/mrm.24916

Huber, L., Kronbichler, L., Stirnberg, R., Ehses, P., Stöcker, T., Fernández-Cabello, S., Poser, B. A., & Kronbichler, M. (2022). Evaluating the capabilities and challenges of layer-fMRI VASO at 3T. *bioRxiv*, 2022.2007.2026.501554. https://doi.org/10.1101/2022.07.26.501554

Huber, L., Poser, B. A., Bandettini, P. A., Arora, K., Wagstyl, K., Cho, S., Goense, J., Nothnagel, N., Morgan, A. T., van den Hurk, J., Muller, A. K., Reynolds, R. C., Glen, D. R., Goebel, R., & Gulban, O. F. (2021a). LayNii: A software suite for layer-fMRI. *NeuroImage, 237*, 118091. https://doi.org/10.1016/j.neuroimage.2021.118091

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. https://doi.org/10.1016/j.neuroimage.2015.11.025

Kano, F., Krupenye, C., Hirata, S., Tomonaga, M., & Call, J. (2019). Great apes use self-experience to anticipate an agent’s action in a false-belief test. *Proceedings of the National Academy of Sciences, 116*(42), 20904-20909. https://doi.org/10.1073/pnas.1910095116

Kernbach, J. M., Yeo, B. T. T., Smallwood, J., Margulies, D. S., Thiebaut de Schotten, M., Walter, H., Sabuncu, M. R., Holmes, A. J., Gramfort, A., Varoquaux, G., Thirion, B., & Bzdok, D. (2018). Subspecialization within default mode nodes characterized in 10,000 UK Biobank participants. *Proceedings of the National Academy of Sciences, 115*(48), 12295-12300. https://doi.org/10.1073/pnas.1804876115

Keysers, C., & Gazzola, V. (2007). Integrating simulation and theory of mind: from self to social cognition. *Trends in Cognitive Sciences, 11*(5), 194-196. https://doi.org/10.1016/j.tics.2007.02.002

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

Kim, E., Kyeong, S., Cheon, K.-A., Park, B., Oh, M.-K., Chun, J. W., Park, H.-J., Kim, J.-J., & Song, D.-H. (2016). Neural responses to affective and cognitive theory of mind in children and adolescents with autism spectrum disorder. *Neuroscience Letters, 621*, 117-125. https://doi.org/10.1016/j.neulet.2016.04.026

Kliemann, D., Young, L., Scholz, J., & Saxe, R. (2008). The influence of prior record on moral judgment. *Neuropsychologia, 46*(12), 2949-2957. https://doi.org/10.1016/j.neuropsychologia.2008.06.010

Klin, A., & Jones, W. (2006). Attributing social and physical meaning to ambiguous visual displays in individuals with higher-functioning autism spectrum disorders. *Brain and Cognition, 61*(1), 40-53. https://doi.org/10.1016/j.bandc.2005.12.016

Kobayashi, C., Glover, G. H., & Temple, E. (2007). Children's and adults' neural bases of verbal and nonverbal 'theory of mind'. *Neuropsychologia, 45*(7), 1522-1532. https://doi.org/10.1016/j.neuropsychologia.2006.11.017

Kok, P., Bains, L. J., van Mourik, T., Norris, D. G., & de Lange, F. P. (2016). Selective Activation of the Deep Layers of the Human Primary Visual Cortex by Top-Down Feedback. *Current Biology, 26*(3), 371-376. https://doi.org/10.1016/j.cub.2015.12.038

Koopmans, P. J., Barth, M., Orzada, S., & Norris, D. G. (2011). Multi-echo fMRI of the cortical laminae in humans at 7T. *NeuroImage, 56*(3), 1276-1285. https://doi.org/10.1016/j.neuroimage.2011.02.042

Korkmaz, B. (2011). Theory of Mind and Neurodevelopmental Disorders of Childhood. *Pediatric Research, 69*(8), 101-108. https://doi.org/10.1203/PDR.0b013e318212c177

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

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in neurosciences, 23*(11), 571-579. https://doi.org/10.1016/s0166-2236(00)01657-x

Lamme, V. A. F., Supèr, H., & Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Current Opinion in Neurobiology, 8*(4), 529-535. https://doi.org/10.1016/S0959-4388(98)80042-1

Lawrence, S. J. D., Norris, D. G., & de Lange, F. P. (2019). Dissociable laminar profiles of concurrent bottom-up and top-down modulation in the human visual cortex. *eLife, 8*, e44422. https://doi.org/10.7554/eLife.44422

Logothetis, N. K., & Wandell, B. A. (2004). Interpreting the BOLD signal. *Annual Review of Physiology, 66*, 735-769. https://doi.org/10.1146/annurev.physiol.66.082602.092845

Lohmann, H., & Tomasello, M. (2003). The Role of Language in the Development of False Belief Understanding: A Training Study. *Child Development, 74*(4), 1130-1144. https://doi.org/10.1111/1467-8624.00597

Lu, H., Golay, X., Pekar, J. J., & van Zijl, P. C. M. (2003). Functional magnetic resonance imaging based on changes in vascular space occupancy. *Magnetic Resonance in Medicine, 50*(2), 263-274. https://doi.org/10.1002/mrm.10519

Lu, H., Hua, J., & van Zijl, P. C. (2013). Noninvasive functional imaging of cerebral blood volume with vascular-space-occupancy (VASO) MRI. *NMR in Biomedicine, 26*(8), 932-948. https://doi.org/10.1002/nbm.2905

Lu, H., & van Zijl, P. C. M. (2012). A review of the development of Vascular-Space-Occupancy (VASO) fMRI. *NeuroImage, 62*(2), 736-742. https://doi.org/10.1016/j.neuroimage.2012.01.013

Maier, A., Adams, G., Aura, C., & Leopold, D. (2010). Distinct Superficial and Deep Laminar Domains of Activity in the Visual Cortex during Rest and Stimulation. *Frontiers in Systems Neuroscience, 4*. https://doi.org/10.3389/fnsys.2010.00031

Mar, R. A. (2011). The Neural Bases of Social Cognition and Story Comprehension. *Annual Review of Psychology, 62*(1), 103-134. https://doi.org/10.1146/annurev-psych-120709-145406

Markov, N. T., Misery, P., Falchier, A., Lamy, C., Vezoli, J., Quilodran, R., Gariel, M., Giroud, P., Ercsey-Ravasz, M., & Pilaz, L. (2011). Weight consistency specifies regularities of macaque cortical networks. *Cerebral Cortex, 21*(6), 1254-1272. https://doi.org/10.1093/cercor/bhq201

Mars, R., Neubert, F.-X., Noonan, M., Sallet, J., Toni, I., & Rushworth, M. (2012). On the relationship between the “default mode network” and the “social brain” [Hypothesis and Theory]. *Frontiers in human neuroscience, 6*. https://doi.org/10.3389/fnhum.2012.00189

Mars, R. B., Sallet, J., Schüffelgen, U., Jbabdi, S., Toni, I., & Rushworth, M. F. S. (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. https://doi.org/10.1093/cercor/bhr268

Martin, A., & Weisberg, J. (2003). Neural foundations for understanding social and mechanical concepts. *Cognitive Neuropsychology, 20*(3-6), 575-587. https://doi.org/10.1080/02643290342000005

Mason, R. A., & Just, M. A. (2009). The Role of the Theory-of-Mind Cortical Network in the Comprehension of Narratives. *Language and Linguistics Compass, 3*(1), 157-174. https://doi.org/10.1111/j.1749-818X.2008.00122.x

Menon, R. S., Ogawa, S., Hu, X., Strupp, J. P., Anderson, P., & Uğurbil, K. (1995). BOLD Based Functional MRI at 4 Tesla Includes a Capillary Bed Contribution: Echo-Planar Imaging Correlates with Previous Optical Imaging Using Intrinsic Signals. *Magnetic Resonance in Medicine, 33*(3), 453-459. https://doi.org/10.1002/mrm.1910330323

Merriam, E. P., & Kay, K. (2022). The need for validation in layer-specific fMRI. https://doi.org/10.31219/osf.io/f9vqc

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*(1), 167-202. https://doi.org/10.1146/annurev.neuro.24.1.167

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and Theory of Mind: Meta-Analysis of the Relation Between Language Ability and False-belief Understanding. *Child Development, 78*(2), 622-646. https://doi.org/10.1111/j.1467-8624.2007.01018.x

Mitchell, J. P. (2007). Activity in Right Temporo-Parietal Junction is Not Selective for Theory-of-Mind. *Cerebral Cortex, 18*(2), 262-271. https://doi.org/10.1093/cercor/bhm051

Muckli, L., De Martino, F., Vizioli, L., Petro, Lucy S., Smith, Fraser W., Ugurbil, K., Goebel, R., & Yacoub, E. (2015). Contextual Feedback to Superficial Layers of V1. *Current Biology, 25*(20), 2690-2695. https://doi.org/10.1016/j.cub.2015.08.057

Norris, D. G., & Polimeni, J. R. (2019). Laminar (f)MRI: A short history and future prospects. *NeuroImage, 197*, 643-649. https://doi.org/10.1016/j.neuroimage.2019.04.082

Nowak, M. A. (2006). Five Rules for the Evolution of Cooperation. *Science, 314*(5805), 1560-1563. https://doi.org/10.1126/science.1133755

Numssen, O., Bzdok, D., & Hartwigsen, G. (2021). Functional specialization within the inferior parietal lobes across cognitive domains. *eLife, 10*, e63591. https://doi.org/10.7554/eLife.63591

Ocklenburg, S., Beste, C., Arning, L., Peterburs, J., & Güntürkün, O. (2014). The ontogenesis of language lateralization and its relation to handedness. *Neuroscience & Biobehavioral Reviews, 43*, 191-198. https://doi.org/10.1016/j.neubiorev.2014.04.008

Paquola, C., Amunts, K., Evans, A., Smallwood, J., & Bernhardt, B. (2022). Closing the mechanistic gap: the value of microarchitecture in understanding cognitive networks. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2022.07.001

Paquola, C., Vos De Wael, R., Wagstyl, K., Bethlehem, R. A. I., Hong, S.-J., Seidlitz, J., Bullmore, E. T., Evans, A. C., Misic, B., Margulies, D. S., Smallwood, J., & Bernhardt, B. C. (2019). Microstructural and functional gradients are increasingly dissociated in transmodal cortices. *PLoS biology, 17*(5), e3000284. https://doi.org/10.1371/journal.pbio.3000284

Paunov, A. M., Blank, I. A., & Fedorenko, E. (2019). Functionally distinct language and Theory of Mind networks are synchronized at rest and during language comprehension. *Journal of Neurophysiology, 121*(4), 1244-1265. https://doi.org/10.1152/jn.00619.2018

Paunov, A. M., Blank, I. A., Jouravlev, O., Mineroff, Z., Gallée, J., & Fedorenko, E. (2022). Differential Tracking of Linguistic vs. Mental State Content in Naturalistic Stimuli by Language and Theory of Mind (ToM) Brain Networks. *Neurobiology of Language, 3*(3), 413-440. https://doi.org/10.1162/nol_a_00071

Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Thinking of mental and other representations: The roles of left and right temporo-parietal junction. *Social Neuroscience, 1*(3-4), 245-258. https://doi.org/10.1080/17470910600989896

Perner, J., & Leekam, S. (2008). The Curious Incident of the Photo that was Accused of Being False: Issues of Domain Specificity in Development, Autism, and Brain Imaging. *Quarterly Journal of Experimental Psychology, 61*(1), 76-89. https://doi.org/10.1080/17470210701508756

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology, 5*(2), 125-137. https://doi.org/10.1111/j.2044-835X.1987.tb01048.x

Pinker, S. (2010). The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences, 107*, 8993-8999. https://doi.org/10.1073/pnas.0914630107

Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Social Cognitive and Affective Neuroscience, 2*(1), 67-70. https://doi.org/10.1093/scan/nsm006

Poletti, M., Enrici, I., & Adenzato, M. (2012). Cognitive and affective Theory of Mind in neurodegenerative diseases: Neuropsychological, neuroanatomical and neurochemical levels. *Neuroscience & Biobehavioral Reviews, 36*(9), 2147-2164. https://doi.org/10.1016/j.neubiorev.2012.07.004

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515-526. https://doi.org/10.1017/S0140525X00076512

Puce, A., & Perrett, D. (2003). Electrophysiology and brain imaging of biological motion. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*(1431), 435-445. https://doi.org/10.1098/rstb.2002.1221

Pyers, J. E., & Senghas, A. (2009). Language promotes false-belief understanding: evidence from learners of a new sign language. *Psychological Science, 20*(7), 805-812. https://doi.org/10.1111/j.1467-9280.2009.02377.x

Raichle, M. E. (2015). The brain's default mode network. *Annual Review of Neuroscience, 38*, 433-447. https://doi.org/10.1146/annurev-neuro-071013-014030

Rauschenberger, R. (2010). Reentrant processing in attentional guidance — Time to abandon old dichotomies. *Acta Psychologica, 135*(2), 109-111. https://doi.org/10.1016/j.actpsy.2010.04.014

Rauss, K., & Pourtois, G. (2013). What is Bottom-Up and What is Top-Down in Predictive Coding? *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00276

Richardson, H., Koster-Hale, J., Caselli, N., Magid, R., Benedict, R., Olson, H., Pyers, J., & Saxe, R. (2020). Reduced neural selectivity for mental states in deaf children with delayed exposure to sign language. *Nature Communications, 11*(1), 3246. https://doi.org/10.1038/s41467-020-17004-y

Richardson, H., Lisandrelli, G., Riobueno-Naylor, A., & Saxe, R. (2018). Development of the social brain from age three to twelve years. *Nature Communications, 9*(1), 1027. https://doi.org/10.1038/s41467-018-03399-2

Richerson, P. J., & Boyd, R. (1998). The evolution of human ultra-sociality. *Indoctrinability, ideology, and warfare: Evolutionary perspectives*, 71-95.

Rockland, K. S., & Pandya, D. N. (1979). Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Research, 179*(1), 3-20. https://doi.org/10.1016/0006-8993(79)90485-2

Roumazeilles, L., Schurz, M., Lojkiewiez, M., Verhagen, L., Schüffelgen, U., Marche, K., Mahmoodi, A., Emberton, A., Simpson, K., Joly, O., Khamassi, M., Rushworth, M. F. S., Mars, R. B., & Sallet, J. (2021). Social prediction modulates activity of macaque superior temporal cortex. *Science Advances, 7*(38), eabh2392. https://doi.org/doi:10.1126/sciadv.abh2392

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in “theory of mind”. *NeuroImage, 19*(4), 1835-1842. https://doi.org/10.1016/S1053-8119(03)00230-1

Saxe, R., & Powell, L. J. (2006). It's the Thought That Counts:Specific Brain Regions for One Component of Theory of Mind. *Psychological Science, 17*(8), 692-699. https://doi.org/10.1111/j.1467-9280.2006.01768.x

Saxe, R., & Wexler, A. (2005). Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia, 43*(10), 1391-1399. https://doi.org/10.1016/j.neuropsychologia.2005.02.013

Saygin, A. P. (2007). Superior temporal and premotor brain areas necessary for biological motion perception. *Brain, 130*(9), 2452-2461. https://doi.org/10.1093/brain/awm162

Schaafsma, S. M., Pfaff, D. W., Spunt, R. P., & Adolphs, R. (2015). Deconstructing and reconstructing theory of mind. *Trends in Cognitive Sciences, 19*(2), 65-72. https://doi.org/10.1016/j.tics.2014.11.007

Schafroth, J. L., Basile, B. M., Martin, A., & Murray, E. A. (2021). No evidence that monkeys attribute mental states to animated shapes in the Heider–Simmel videos. *Scientific Reports, 11*(1), 3050. https://doi.org/10.1038/s41598-021-82702-6

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Psychology Press.

Scheeringa, R., Koopmans, P. J., van Mourik, T., Jensen, O., & Norris, D. G. (2016). The relationship between oscillatory EEG activity and the laminar-specific BOLD signal. *Proceedings of the National Academy of Sciences, 113*(24), 6761-6766. https://doi.org/10.1073/pnas.1522577113

Schilbach, L., Eickhoff, S. B., Rotarska-Jagiela, A., Fink, G. R., & Vogeley, K. (2008). Minds at rest? Social cognition as the default mode of cognizing and its putative relationship to the "default system" of the brain. *Consciousness and Cognition, 17*(2), 457-467. https://doi.org/10.1016/j.concog.2008.03.013

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences, 4*(8), 299-309. https://doi.org/10.1016/S1364-6613(00)01506-0

Scholz, J., Triantafyllou, C., Whitfield-Gabrieli, S., Brown, E. N., & Saxe, R. (2009). Distinct Regions of Right Temporo-Parietal Junction Are Selective for Theory of Mind and Exogenous Attention. *PLOS ONE, 4*(3), e4869. https://doi.org/10.1371/journal.pone.0004869

Schurz, M., & Perner, J. (2015). An evaluation of neurocognitive models of theory of mind. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.01610

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. https://doi.org/10.1016/j.neubiorev.2014.01.009

Schurz, M., Radua, J., Tholen, M. G., Maliske, L., Margulies, D. S., Mars, R. B., Sallet, J., & Kanske, P. (2021). Toward a hierarchical model of social cognition: A neuroimaging meta-analysis and integrative review of empathy and theory of mind. *Psychological bulletin*, 293-327. https://doi.org/10.1037/bul0000303

Seghier, M. L. (2013). The Angular Gyrus: Multiple Functions and Multiple Subdivisions. *The Neuroscientist, 19*(1), 43-61. https://doi.org/10.1177/1073858412440596

Self, M. W., van Kerkoerle, T., Supèr, H., & Roelfsema, P. R. (2013). Distinct roles of the cortical layers of area V1 in figure-ground segregation. *Curr Biol, 23*(21), 2121-2129. https://doi.org/10.1016/j.cub.2013.09.013

Shain, C., Paunov, A., Chen, X., Lipkin, B., & Fedorenko, E. (2022). No evidence of theory of mind reasoning in the human language network. *bioRxiv*, 2022.2007.2018.500516. https://doi.org/10.1101/2022.07.18.500516

Silver, M. S., Joseph, R. I., & Hoult, D. I. (1984). Highly selective π2 and π pulse generation. *Journal of Magnetic Resonance (1969), 59*(2), 347-351. https://doi.org/10.1016/0022-2364(84)90181-1

Sohn, P., & Reher, K. (2009). *Partly cloudy.* Walt Disney Pictures. Pixar Animation Studios.

Sommer, M., Döhnel, K., Sodian, B., Meinhardt, J., Thoermer, C., & Hajak, G. (2007). Neural correlates of true and false belief reasoning. *NeuroImage, 35*(3), 1378-1384. https://doi.org/10.1016/j.neuroimage.2007.01.042

Squire, L. R., & Zola-Morgan, S. (1991). The Medial Temporal Lobe Memory System. *Science, 253*(5026), 1380-1386. https://doi.org/10.1126/science.1896849

Stirnberg, R., & Stöcker, T. (2021). Segmented K-space blipped-controlled aliasing in parallel imaging for high spatiotemporal resolution EPI. *Magnetic Resonance in Medicine, 85*(3), 1540-1551. https://doi.org/10.1002/mrm.28486

Stuss, D. T., Gallup, G. G., Jr., & Alexander, M. P. (2001). The frontal lobes are necessary for `theory of mind'. *Brain, 124*(2), 279-286. https://doi.org/10.1093/brain/124.2.279

Tannús, A., & Garwood, M. (1997). Adiabatic pulses. *NMR in Biomedicine, 10*(8), 423-434.

Turner, R. (2002). How Much Cortex Can a Vein Drain? Downstream Dilution of Activation-Related Cerebral Blood Oxygenation Changes. *NeuroImage, 16*(4), 1062-1067. https://doi.org/10.1006/nimg.2002.1082

Uludağ, K., & Blinder, P. (2018). Linking brain vascular physiology to hemodynamic response in ultra-high field MRI. *NeuroImage, 168*, 279-295. https://doi.org/10.1016/j.neuroimage.2017.02.063

Van Essen, D. C., & Maunsell, J. H. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends in neurosciences, 6*, 370-375. https://doi.org/10.1016/0166-2236(83)90167-4

Van Overwalle, F., & Vandekerckhove, M. (2013). Implicit and explicit social mentalizing: dual processes driven by a shared neural network. *Frontiers in human neuroscience, 7*. https://doi.org/10.3389/fnhum.2013.00560

van Veluw, S. J., & Chance, S. A. (2014). Differentiating between self and others: an ALE meta-analysis of fMRI studies of self-recognition and theory of mind. *Brain Imaging and Behavior, 8*(1), 24-38. https://doi.org/10.1007/s11682-013-9266-8

Vandewouw, M. M., Safar, K., Mossad, S. I., Lu, J., Lerch, J. P., Anagnostou, E., & Taylor, M. J. (2021). Do shapes have feelings? Social attribution in children with autism spectrum disorder and attention-deficit/hyperactivity disorder. *Translational Psychiatry, 11*(1), 493. https://doi.org/10.1038/s41398-021-01625-y

Varley, R., & Siegal, M. (2000). Evidence for cognition without grammar from causal reasoning and ‘theory of mind’ in an agrammatic aphasic patient. *Current Biology, 10*(12), 723-726. https://doi.org/10.1016/S0960-9822(00)00538-8

Varley, R., Siegal, M., & Want, S. C. (2001). Severe Impairment in Grammar Does Not Preclude Theory of Mind. *Neurocase, 7*(6), 489-493. https://doi.org/10.1093/neucas/7.6.489

Vu, A. T., Jamison, K., Glasser, M. F., Smith, S. M., Coalson, T., Moeller, S., Auerbach, E. J., Uğurbil, K., & Yacoub, E. (2017). Tradeoffs in pushing the spatial resolution of fMRI for the 7T Human Connectome Project. *NeuroImage, 154*, 23-32. https://doi.org/10.1016/j.neuroimage.2016.11.049

Wagstyl, K., Larocque, S., Cucurull, G., Lepage, C., Cohen, J. P., Bludau, S., Palomero-Gallagher, N., Lewis, L. B., Funck, T., Spitzer, H., Dickscheid, T., Fletcher, P. C., Romero, A., Zilles, K., Amunts, K., Bengio, Y., & Evans, A. C. (2020). BigBrain 3D atlas of cortical layers: Cortical and laminar thickness gradients diverge in sensory and motor cortices. *PLoS biology, 18*(4), e3000678. https://doi.org/10.1371/journal.pbio.3000678

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development, 72*(3), 655-684. https://doi.org/10.1111/1467-8624.00304

Willems, R. M., Benn, Y., Hagoort, P., Toni, I., & Varley, R. (2011). Communicating without a functioning language system: Implications for the role of language in mentalizing. *Neuropsychologia, 49*(11), 3130-3135. https://doi.org/10.1016/j.neuropsychologia.2011.07.023

Wilson, A. C. (2021). Do animated triangles reveal a marked difficulty among autistic people with reading minds? *Autism, 25*(5), 1175-1186. https://doi.org/10.1177/1362361321989152

Wimmer, H., Hogrefe, G.-J., & Perner, J. (1988). Children's Understanding of Informational Access as Source of Knowledge. *Child Development, 59*(2), 386-396. https://doi.org/10.2307/1130318

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103-128. https://doi.org/10.1016/0010-0277(83)90004-5

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature methods, 8*(8), 665-670. https://doi.org/10.1038/nmeth.1635

Yoshida, W., Dolan, R. J., & Friston, K. J. (2008). Game Theory of Mind. *PLOS Computational Biology, 4*(12), e1000254. https://doi.org/10.1371/journal.pcbi.1000254

Young, L., Camprodon, J. A., Hauser, M., Pascual-Leone, A., & Saxe, R. (2010). Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proceedings of the National Academy of Sciences, 107*(15), 6753-6758. https://doi.org/10.1073/pnas.0914826107

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences, 104*(20), 8235-8240. https://doi.org/10.1073/pnas.0701408104

Zahn, R., Moll, J., Krueger, F., Huey, E. D., Garrido, G., & Grafman, J. (2007). Social concepts are represented in the superior anterior temporal cortex. *Proceedings of the National Academy of Sciences, 104*(15), 6430-6435. https://doi.org/10.1073/pnas.0607061104

Zaitchik, D. (1990). When representations conflict with reality: The preschooler's problem with false beliefs and “false” photographs. *Cognition, 35*(1), 41-68. https://doi.org/10.1016/0010-0277(90)90036-J

Appendix

The original scripts that served as the basis for further customizations can be found here:

https://github.com/layerfMRI/repository/tree/master/3T_VASO_scripts

Supplementary Table 1

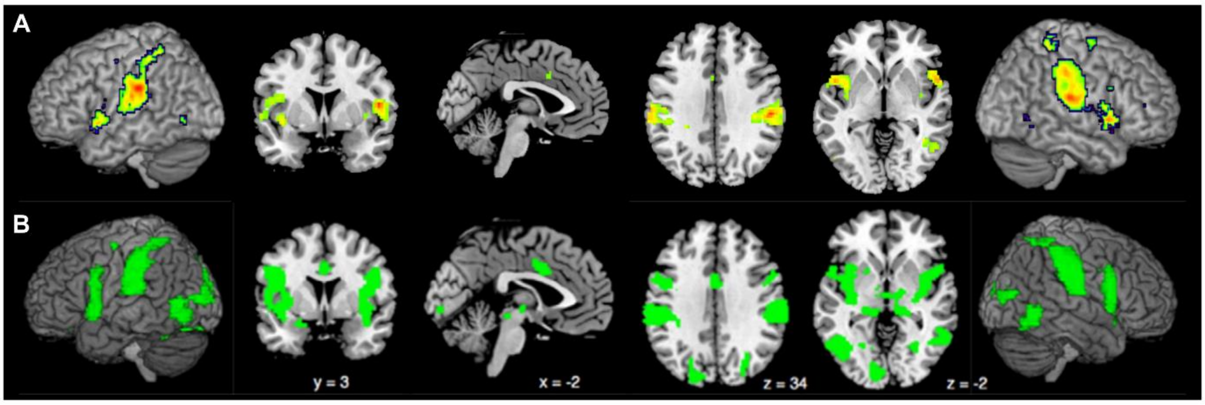
Significant Clusters for the "Pain > Belief" Contrast of the Functional Localizer

Cluster	Region	Nr. of voxels	x	y	z	Peak *t*
1	*Left Supramarginal Gyrus*	782	-57	-30	28	13.64
2	*Left Middle Cingulate Gyrus*	291	50	3	12	12.89
3	*Right Supramarginal Gyrus*	849	53	-24	35	12.71
4	*Right Superior Parietal Lobule*	245	26	-42	55	11.15
5	*Left Cuneus*	357	-50	8	-2	10.88
6	*Right Fusiform Gyrus*	82	40	-52	0	8.59
7	*Left Exterior Cerebellum*	50	-24	-64	-20	8.57
8	*Right Superior Frontal Gyrus*	32	28	-4	65	7.49

Note. Significant clusters (p_{FWE} = .05, *k* = 20). Labels of brain regions obtained from the *Neuromorphometrics* atlas. Peak coordinates of local clusters in MNI space.

Supplementary Figure 1

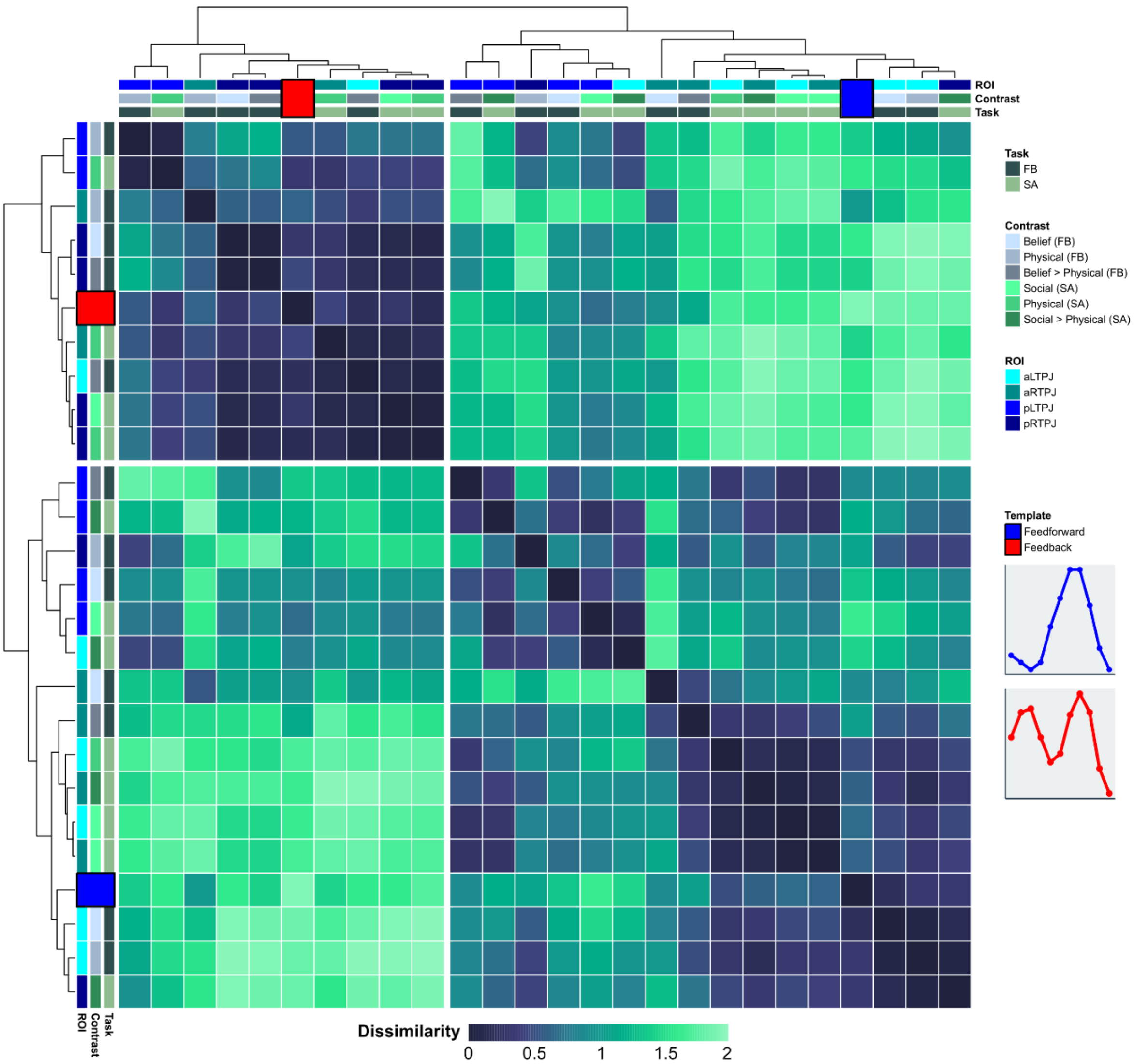
Group-Level Activation Pattern for the “Pain > Belief” Contrast of the Functional Localizer



Note. Whole-brain responses with significant clusters (p_{FWE} < .05, *k* = 0) based on the “*Pain* > *Belief*” contrast of the movie watching task. **(A)** The obtained results in the present experiment show activations in bilateral *Supramarginal Gyrus*, *Left Middle Cingulate Gyrus*, *Right Superior Parietal Lobule*, *Left Cuneus*, and other regions. **(B)** Original results adopted from Jacoby et al. (2016) as a gold-standard comparison for the functional localizer.

Supplementary Figure 2

Complete Hierarchical Cluster Analysis of Layer Profiles



Note. Complete hierarchical cluster analysis of layer profiles including feedforward and feedback templates adapted from Huber et al. (2021b).