The Effects of Self-Relevance on the Modulation of Brain Responses to Social Stimuli

by

Taylor D. Guthrie

A dissertation accepted and approved in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Social Cognitive Neuroscience

Dissertation Committee:

Robert S. Chavez, Chair

Elliot Berkman, Core Member

Jennifer Pfeifer, Core Member

Nicole Guiliani, Institutional Representative

University of Oregon

Fall 2024

DISSERTATION ABSTRACT

Taylor D. Guthrie

Doctor of Philosophy in Social Cognitive Neuroscience

Title: The Effects of Self-Relevance on the Modulation of Brain Responses to Social Stimuli

This dissertation investigates the neural mechanisms underlying self-referential processing and its influence on cognitive resource allocation and social cognition. Through a series of fMRI studies employing novel experimental designs and multivariate analysis techniques we provide evidence for the self as a powerful cognitive construct that shapes information processing across multiple levels of brain function. We first demonstrate that self-relevant narratives elicit widespread activation and increased neural synchrony across cortical hierarchies. This is seen from low-level sensory regions to higher-order cognitive areas and highlight the self's ability to modulate attention and enhance processing of personally significant information. The second study reveals that social relationship strength predicts the uniqueness of neural representations in key social cognition regions. Closer relationships are found to be associated with more individuated patterns in the ventral medial prefrontal cortex (vmPFC) and anterior insula. Surprisingly, posterior regions like the posterior cingulate cortex (PCC) show more generalized representations for close others which align with behavioral findings of normative trait endorsements. Lastly, we successfully decode person identity from neural activity patterns in default mode network regions which demonstrate the presence of reliable social information that can differentiate familiar individuals. These findings advance our understanding of how the self influences cognitive processing and shapes our perceptions of others. This work contributes to a more mechanistic account of self-referential processing and its role in guiding social cognition.

This dissertation includes previously published coauthored material

ACKNOWLEDGMENTS

First and foremost, I would like to extend my deepest gratitude to Dr. Rob Chavez for his invaluable guidance throughout this journey. His expertise in designing and funding the experiments, as well as his mentorship in data analysis and manuscript preparation, has been instrumental in the success of this work. I am deeply grateful for his unwavering support and commitment. I would also like to thank Jack Kapustka for his exceptional managerial contributions. His dedication to overseeing data collection and ensuring the smooth progress of our projects made this work possible. A special mention goes to Austin Frost and Nate Holley, whose computational contributions were crucial to the success of this dissertation. Their efforts in setting up the processing pipelines for data analysis were instrumental and greatly appreciated. Additionally, I would like to acknowledge all the research assistants who have volunteered their time and effort in the Computational Social Neuroscience Lab (CSNL) over the years. Without their dedication and hard work, none of this would have been possible. Thank you all for your contributions, guidance, and support!

DEDICATION

This dissertation is dedicated to my spouse, Emma, whose love and support were invaluable as I navigated this challenging journey. Your constant encouragement to follow my passion and creativity has been my driving force and I am forever grateful for the strength and motivation you provided me. To my children, Thor and Dixon, you are the reason I strive to be the best scientist and researcher I can be. My hope is that one day you will look back with pride on what your father has accomplished. I also dedicate this work to my mother and father, for their financial and moral support, and to my brothers and sisters, who have always believed in me and helped me realize my potential.

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

Chapter I
General Introduction

Despite the constant barrage of sensory information that the brain receives in every waking moment, it somehow manages to prioritize the processing of information that is relevant to an individual's goals, experiences, and preferences (Beer et al., 2010; Heatherton, 2011; Meyer & Lieberman, 2018; Northoff et al., 2006; Northoff & Hayes, 2011). This prioritization likely revolves around a core process of selfhood, a multifaceted cognitive construct that seems to have a powerful influence on the allocation of cognitive resources and the coordination of various processing streams. (Sui et al., 2013; Sui & Humphreys, 2015a). This sense of self is likely shaped by the pressures an individual faces and the needs that they must work to fulfill (Pfeifer & Berkman, 2018). As a highly social species, this need space includes the management of social relationships with others as they are paramount to survival in a social context (Aron et al., 1991; Heatherton et al., 2006; Mitchell, Banaji, et al., 2005a; Wagner et al., 2012). This dissertation aims to investigate the neural mechanisms underlying the self's ability to influence cognitive processing and to explore how the brain differentially processes information related to the self and others, particularly those with whom we have close social connections.

The idea of selfhood and introspection has been a central topic of philosophical inquiry from early Greek writings to modern day discourse. The modern neuroscience exploration into the mechanistic underpinnings of self-related processing all stem from the theoretical foundations laid by these early contemporaries that were later elaborated on and tested by early cognitive scientists. However, recent advancements in neuroimaging techniques and computational modeling approaches (Haxby, 2012; Kriegeskorte, 2011) have only recently opened the door for a comprehensive test of these theoretical frameworks in accordance with

actual brain processes. By leveraging these new methodological approaches, this work seeks to elucidate how self-related processing modulates the allocation of cognitive resources and coordinates various stimulus processing streams. Through a series of studies investigating the neural processing of self-relevant and socially-relevant information, this dissertation aims to provide evidence for the self as a powerful cognitive construct that recruits and coordinates cognitive processing streams, such as memory, attention and cognitive control, to deeply process and modify self-related information.

The Philosophical Self

Some of the earliest philosophical writings emphasized the importance of self-reflection and the central role these processes had in maintaining a coherent and meaningful sense of self (Aristotle, 1999; Plato & Segal, 1986). These introspective processes, involved in the examination of our thoughts, feelings and experiences, were believed to be crucial cognitive ingredients that separated our species from other animals and allowed us to build elaborate and meaningful self-concepts and social knowledge. From Socrates' challenge to, "Know thyself," (Plato & Segal, 1986) to Descartes' notion that the act of thinking itself was the only proof of one's existence, philosophers have continuously placed self-reflection at the center of their inquiries into the nature of the self (Descartes & Cottingham, 2013). We build narratives over the course of a lifetime that are a culmination of momentary reflections on our mental state, our beliefs and our values (James, 1890). We use these reflections to challenge ourselves to move past and learn from adversity, to grow, to regulate and to change (Taylor, 1989). This theoretical basis of self-reflection as a basis for self-hood laid the foundation for a concept of self that was seemingly paradoxical, being something that was both continuous and stable but also dynamic and in constant flux. (Erikson, 1994; Mead, 1934).

Much of the philosophical discourse on the self emphasizes the self as an enduring stable core of our being that persists through time and change. John Locke, for instance, argued that memory was the necessary ingredient for having and maintaining a sense of self (Locke, 2000). Without the ability to reflect on our past experiences we would not be able to comprehend our place in the present moment or think about how to navigate the future. William James elaborated on this idea when he introduced the concept of the "I" and "me" selves. The subjective, experiencing, "I" self, seems to be stable and continuous in the face of change providing a sense of unity from one moment to the next whereas the "me" self is reflective and dynamically built in the context of the moment (James, 1890). This all relates to the cognitive process of memory, in which experiences shape a set of core beliefs and preferences that add a filter on to the way in which we see and perceive the world.

This stable self was routinely tied to passive memory processes in these early investigations of the self, whereas the dynamic changeable aspects of the self were described in a more active and engaged manner. The existentialist movement in philosophy emphasized that the self was not a pre-existing essence or soul, but rather that it was something that was consistently being created and managed through an individual's choices and behaviors (Beauvoir et al., 1989; Sartre & Sartre, 2012). The self was described as an ongoing project that was always in the process of becoming, continuously being shaped by the individual's ongoing engagement with the world. These philosophers highlighted the importance of agency and choice in shaping the self, as individuals actively pursued their goals and defined a sense of meaning and belonging in their world. This dynamic self was described as an active process that was capable of utilizing those past stable experiences but in the service of defining a new path forward. Together, these purely reasoned views of the continuous and dynamic elements of selfhood were pointing

towards a cognitive construct that involved neural resources required for memory processing as well as all of those required for the cognitive processing streams involved in motivation, goal pursuit and schema modification.

The highly social elements of human existence contribute to additional nuance when considering the theoretical underpinnings of selfhood that arose from the philosophical discourse. Some philosophers have gone as far as saying that the self exists only because of our extensive social interactions with others as we internalize the attitudes and perspectives that others have of us (Mead, 1934). Feminist philosophers such as Simone de Beauvoir and Judith Butler strongly emphasize the extraordinary impact that social norms, roles and expectations have on our concept of self and the beliefs we develop about our place in the world (Beauvoir et al., 1989; Butler, 2006). They argue that the self is a product of an ongoing process of "performativity," where individuals enact and embody social scripts and identities that are encouraged by their culture and society. While these philosophical perspectives highlight the inherently social nature of the human self, it can also be argued that self does not wholly rely on sociality for its existence. It could be that these social relationships are just highly valuable to us as humans and are part of a broader connection of selfhood to the goal pursuit and motivation processes discussed earlier (Maslow, 1943).

These various lines of philosophical insight all point to a concept of self that is continuous and stable over time, is dynamic and capable of change and growth and is essential for social navigation and interaction. These theories all suggest that the self is a powerful cognitive construct that is a central organizing principle in human cognition. Kant captured this idea with his notion of the "transcendental unity of apperception" which proposed that the self

plays a significant role in organizing all of the varied components of cognition, from sensory input and memories to thoughts and beliefs (Kant et al., 2000). It paints the theoretical self as a golden thread that all other cognitive processes are tied to, strongly influencing the way that information about the world is prioritized and processed. James highlighted this when discussing the role of selective attention, the ability to focus solely on a narrow range of the incoming sensory information, as a prioritization process that is highly self-relevant in nature (James, 1890). These perspectives underscore the idea that the self is deeply embedded in the core architecture of human cognition, serving as a central hub that selects, integrates and deeply processes information that is relevant to the goals of the individual.

Cognitive Investigations of the Self

The theoretical foundations of the self that philosophers arrived at through reason and introspection alone were eventually put to the test with the rise of cognitive science. This transition allowed for a translation of hundreds of years of insights into testable hypotheses and experimental paradigms. The cognitive perspective was crucial as it allowed for the integration of self-related processes with other areas of cognition being studied at the time, such as memory, attention, and cognitive control. By providing a more comprehensive understanding of how self-related stimuli were able to influence these processes, this empirical approach allowed for the construction of a foundational basis of cognitive mechanisms that give rise to the self. It became increasingly clear through behavioral empirical support that the self was indeed a multifaceted construct that plays a central role in organizing and guiding information processing.

The early cognitive studies on the self focused on the relationship between self-reference and memory. A seminal finding in this area was the discovery of the self-reference effect by Rogers, Kuiper, and Kirker (1977). The self-reference effect refers to the phenomenon whereby

information that is processed in relation to the self is better remembered than information processed in other ways. In their study, participants were asked to rate adjectives on various dimensions, including self-referential ("Describes me?"), semantic ("Means the same as xxx?"), and structural ("Big letters?") tasks. The results showed a distinct and robust memory advantage for the adjectives rated in the self-referential task which were recalled better than those rated in the other tasks.

This finding sparked a debate about the nature of the self-reference effect and its implications for the self as a cognitive construct. One perspective, championed by Bower and Gilligan (1979) and Kuiper and Rogers (1979), argued that the self-reference effect could be explained by the depth of processing involved in self-referential encoding, and suggested that the self was powerful but ordinary (Greenwald & Banaji, 1989). They suggested that processing information in relation to the self leads to a deeper, more elaborate encoding, resulting in better memory performance. In contrast, the competing view, put forward by Klein and Kihlstrom (1986) and Klein and Loftus (1988), proposed that the self is a unique cognitive structure with special mnemonic properties. According to this perspective, the self-schema is not merely a deep level of processing but a distinct and highly organized knowledge structure that facilitates the encoding and retrieval of self-related information. Unfortunately, with the tools available at the time there was no way to determine which of these theories was correct due to them both making the same behavioral predictions that words that were self-relevant would be associated with higher memory performance.

This work set the stage for the creation of theoretical frameworks that attempted to describe the self as a powerful and unique cognitive construct that is treated differently in the

brain then schemas that are more semantic in nature. The self-schema theory put forth by Hazel Markus (1977) emerged as one of the most prominent frameworks that attempted to describe the role that the self has in information processing and behavior. The theory proposed a concept of the self as a schematic knowledge structure that is intimately involved in the organization and prioritization of incoming sensory information. This web of knowledge associations was believed to be comprised of an individual's understanding of their own traits, abilities, preferences, and experiences.

The self-schema theory proposed that the self-construct was actively involved in various cognitive processes such as attention, perception, memory, and inference, allowing for self-related information to be engaged cognitively in a different way than unrelated semantic information (Markus, 1977). Markus argued, in line with philosophical foundations, that this construct had to be dynamic in nature and capable of undergoing change and growth as it encountered information that challenged core beliefs stored within the schema. Social interaction and engagement with the social world were believed to be a crucial component to the change process as these experiences provide ample self-relevant feedback (Markus, 1977). This describes the self as a persistent, socially embedded, knowledge structure that is not just accessed when the need arises, like other semantic associative structures, but rather a central organizing construct that guides the filtering of information for the purposes of dynamic remodeling.

Many lines of behavioral research have provided empirical support for the pervasive influence of self-schemas on cognitive processing that the self-schema theory describes. Individuals who believe that a particular trait is included in their self-schema exhibit faster processing and better memory for congruent trait information in comparison to traits that are

incongruent with an individual's self-schema. (Markus, 1977; Markus et al., 1982). This demonstrates that these effects extend beyond memory, with studies showing heighted reaction time for self-referent information than non-self-referent information (Kuiper & Rogers, 1979; Rogers et al., 1977). Moreover, self-schemas have been shown to guide attention, with individuals displaying heightened sensitivity to schema-consistent information (Cherry, 1953; H. Markus et al., 1985, 1987). The robustness of these findings across different cognitive domains and experimental paradigms underscores the central role of self-schemas in organizing and structuring self-related information. The memory advantages, along with the self-schema's ability to guide attention and accelerate information processing, suggest that the cognitive structure is not just a passive repository of self-knowledge but an active mechanism that shapes our perceptions, thoughts, and behaviors (Markus & Kitayama, 1991).

These behavioral studies provided a strong foundation of empirical evidence for the self as an integrated and powerful cognitive construct, but the lack of direct access to the neural processes underlying self-related processes limited the ability of researchers to fully test and refine their models. The debates around whether the self was a unique neural process or merely a deeper depth of processing, like semantic cognition, could not be resolved without direct access to the neural substrate. The advent of neuroimaging techniques in the 1990s, particularly positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), marked a turning point in the empirical study of self-related processing. The ability to observe brain function allowed researchers to test and refine these cognitive and philosophical theories and develop new models that integrated cognitive and neural levels of analysis.

Cognitive Neuroscience of the Self

With the advent of neuroimaging techniques, Craik et al. (1999) employed positron emission tomography (PET) to examine the brain regions activated during self-referential encoding and revealed increased activity in the left frontal cortex and anterior cingulate cortex (ACC). This study provided the first neural evidence for the distinct correlates of self-referential processing. Kelley et al. (2002) built upon these findings by utilizing an event-related functional magnetic resonance imaging (fMRI) paradigm to compare the neural activity associated with self-referential judgments to that of other-referential and case judgments. Their results demonstrated that self-referential processing was more strongly associated with increased activation in the medial prefrontal cortex (mPFC) and posterior cingulate cortex (PCC) compared to processing that was purely semantic in nature. These seminal studies provided evidence for the unique neural underpinnings of self-referential processing and supported the cognitive theory that the self is a distinct cognitive structure and not merely another form of semantic processing.

These initial findings sparked a surge of interest in the neural bases of self-referential processing and the brain regions and networks involved in self-related cognition. A meta-analysis by Northoff et al. (2006) revealed consistent activation in the cortical midline structures, including the mPFC, ACC, and PCC, during self-referential tasks. These cortical midline structures regions have been increasingly recognized as key components of the default mode network (Buckner et al., 2008; Raichle et al., 2001), a large-scale brain network implicated in various self-related processes, such as autobiographical memory retrieval, future planning, and self-reflection (Andrews-Hanna et al., 2014; Spreng et al., 2009). The overlap between the cortical midline structures and the default mode network suggests that self-referential processing may be a fundamental function of this network (Qin & Northoff, 2011; Whitfield-Gabrieli et al., 2011). Further research has explored the specific roles of the default mode network subsystems

in self-related cognition, with the ventral medial prefrontal cortex (vmPFC) and the PCC forming a "midline core" subsystem involved in the processing of self-relevant information and the integration of interoceptive and exteroceptive signals (Andrews-Hanna et al., 2010).

Self-referential processing has been shown to involve a wide range of features from other cognitive domains. Memory is a core component of the theoretical basis of the self-concept and a large body of work has recognized the importance of these regions of the default mode network in facilitating the access and use of autobiographical memory (Buckner & Carroll, 2007; Spreng & Grady, 2010; Svoboda et al., 2006). The vmPFC and PCC have been consistently implicated in these memory processes and have also been shown to coordinate their activity with control network regions such as the dorsolateral prefrontal cortex when self-memory is being utilized to solve problems (Spreng et al., 2009, 2010). Similar to these findings, research has also demonstrated that these same regions along with other nodes in the default mode network are involved in future thinking and prospection, or the ability to imagine and plan for future events (Buckner & Carroll, 2007; D'Argembeau et al., 2014; Schacter et al., 2007). This mental time travel is a fundamental human ability that allows for long-term goal setting, a crucial component of identity formation and maintenance (Bluck & Alea, 2009; Suddendorf & Corballis, 2007).

These long-term goals and preferences, formed through this prospective goal setting process, likely serve as the basis for passive spontaneous thought, which engages the vmPFC and other regions of the default mode network (Andrews-Hanna et al., 2014; Christoff et al., 2009; D'Argembeau et al., 2011; Spreng et al., 2010). This is supported by work showing that the content of spontaneous thought tends to be self-reflective in nature and often involves personal past experiences or future plans (Andrews-Hanna et al., 2013, 2014; Christoff et al., 2009;

Smallwood & Schooler, 2015). Meyer and Lieberman (2018) showed that mPFC activity during rest primes self-referential processing, suggesting that the brain may default to self-reflective states. This is likely adaptive in nature as these thoughts allow one to prepare for future events, maintain a sense of identity and continuity of self, and navigate the complexities of social engagement (Andrews-Hanna et al., 2014).

An important aspect of this self-referential activity is that it tends to be skewed towards information that is positively biased or self-enhancing (Alicke & Sedikides, 2009). This suggests a mechanism that serves to protect the stable components of the self-schema from information that may be damaging to an individual's goals or self-perception (Hughes & Beer, 2013). Researchers have identified particular regions of the default mode network such as the orbitofrontal cortex (OFC) and the dorsal anterior cingulate cortex (dACC) that show decreased activity as people's propensity to rate themselves more favorably than their peers increase (Beer & Hughes, 2010). Furthermore, the vmPFC seems to be preferentially active for traits that are seen as having high personal value (D'Argembeau et al., 2012). Work by Chavez & Heatherton (2015) elaborated on these ideas by showing that individuals with high self-esteem tend to have greater structural frontostriatal connectivity, showing a link between areas involved in self-referential processing and areas known to be involved in value and motivated cognition.

This link between value and self-related processing is particularly interesting when work, outside of the field of social cognition, regarding value-based choice is considered. The vmPFC, often linked to self-referential cognition, has also been shown to be involved in value computations and in the integration of value-based stimulus properties across a wide range of domains (Bartra et al., 2013; Berkman et al., 2017; Hare et al., 2010; Levy & Glimcher, 2012;

Rangel et al., 2008; Roy et al., 2012). These value signals have been suggested to be universal in terms of stimulus identity, meaning that they represent value for any type of stimulus that may be important for decision making purposes (Lebreton et al., 2009; Padoa-Schioppa & Assad, 2006; Philiastides et al., 2010). The identity-value model was put forth by Berkman et al. (2017) to try to explain these links seen between core identity features and their impact on self-regulation and choice. The identity-value model suggests that core identity features are intrinsically linked to subjective value and that the vmPFC's role in both self-referential processing and value-based decision making may reflect a fundamental integration of identity and value in the brain (Northoff & Hayes, 2011).

This value integration is likely facilitated by attentional mechanisms that gate access of relevant stimuli to higher order processing regions such as the vmPFC through feedback loops between these systems (Lim et al., 2011; Sui et al., 2013; Sui & Humphreys, 2015a). The self-attention network (SAN) theory put forth by Humphreys and Sui (2016) suggests intimate connections between the vmPFC and regions implicated in attentional control such as the left posterior superior temporal sulcus (Allison et al., 2000; DiQuattro & Geng, 2011; R. Saxe & Kanwisher, 2003). Dynamic causal modeling has suggested that there seems to be heightened top-down control from vmPFC to posterior superior temporal sulcus to prime the attentional system to be sensitive to self-related information (Chaumon et al., 2014). Sui et al. (2013) demonstrated consistent heightened activity in both vmPFC and left posterior superior temporal sulcus for simple stimuli that had been related specifically to the self and not to others.

It could be that the vmPFC has emerged in human cognition as a hub for tracking and making meaning out of self-relevant information that is pertinent to long term survival goals

(Roy et al., 2012; Schneider & Koenigs, 2017). Survival, however, is an abstract concept in human life and can refer to physiological, social or even self-concept survival as the need space for humans is much richer than that of other animals who spend most of their lives fulfilling only their physiological needs (Maslow, 1943). As we develop, different aspects of this need space become more salient in accordance with the contextual pressures that we face (Erikson, 1994). During adolescence increased activation was observed in the ventromedial prefrontal cortex for self vs other evaluations as individuals increased in age and progressed through puberty (Cosme et al., 2021; Pfeifer et al., 2013). This increased activity however was tightly linked to social rather than academic evaluations reflecting the increased prioritization in adolescence on social belonging and acceptance from peers.

Adolescence is likely a time when the self-schema is being generated as it seems as though self-appraisals during childhood are being actively created rather than efficiently retrieved. This idea is supported by neuroimaging studies comparing children, adolescents, and adults during self-appraisal tasks which found that children and adolescents show greater activation in the anterior rostral and dmPFC and regions of the ACC compared to adults when making direct self-appraisals (Pfeifer et al., 2007, 2009). This may be evidence that younger individuals are engaging in a more active processes of integration and abstraction of self-knowledge as they navigate new environments that fall outside the protection of dependent relationships (Crone & Dahl, 2012; Pfeifer & Peake, 2012; Sebastian et al., 2008). This neural evidence aligns with Erikson's psychosocial theory of development that describes adolescence as a critical period for identity formation. He theorized that adolescents are actively exploring various roles, values, and beliefs in an active and engaged way to construct a coherent sense of self that will guide them into adulthood (Erikson, 1994).

Recent years have seen an increased application of Multivariate Pattern Analysis (MVPA) techniques to investigate the neural underpinnings of self-referential processing and these studies have provided converging evidence for the role of the vmPFC in self and value related processing. Chavez et al. (2017) demonstrated that a classifier trained to distinguish between positively and negatively valanced images could also discriminate between self-referential and other-referential processing in the vMPFC. This went beyond simply looking at increased activation for a contrast between these conditions and instead showed that there were patterns of activity within these regions that seemed to represent both positive affect and self-concept. Similar to these findings, Yankouskaya et al. (2017) showed that patterns of vMPFC activity associated with high-value stimuli could classify self-related information as well but showed that the strongest accuracy was in anterior portions of this region and decreased as the region of interest was moved more posteriorly. More recent work, using RSA, has also demonstrated that the activity in the mPFC is likely representing self-importance rather than just self-descriptiveness, further highlighting the representational role of this region in linking identity to personal value (Levorsen et al., 2023).

This work cumulatively describes a neural mechanism for self-related processes that is centered around the vmPFC, but one that includes the coordination of many large-scale cognitive systems like those involved in memory (Spreng & Grady, 2010), value and decision making (Berkman et al., 2017; Hare et al., 2010), and attention (Humphreys & Sui, 2016). This fits with the description of self as a stable entity on one hand but also something that is built and adjusted throughout the developmental process. It suggests the existence of a cybernetic like self-construct (Syed et al., 2019; Wiener, 2007), one that is continuously monitoring for any information relevant to ongoing goals or perceptions of the self that has the ability to influence a

change in behavior to align incoming information with expected models of identity. This work has provided a grounded mechanistic framework for the existence of the theoretical self-schema that can powerfully shape the allocation of cognitive resources to prioritize the processing of self-related information.

The Self in Social Contexts

While it has been shown that the default mode network is highly involved in self-related processes, it has also been demonstrated to be involved in various other aspects of social cognition such as person perception and its related processes (Mars et al., 2012; Schilbach et al., 2008; Wagner et al., 2012). Much of the work that has been done regarding self-referential processing has been in the form of contrasts between activity that is heightened for self vs that for a particular other (Chavez, 2021; Kelley et al., 2002; Mitchell, Banaji, et al., 2005a; Wagner et al., 2012), demonstrating dissociations between the two within the broader architecture of the default mode network. This distinction is especially prominent within the mPFC such that thinking of others tends to be associated with heightened activity in the dmPFC whereas reflecting on the self tends to be more involved with activation in the vmPFC (Heatherton, 2011; Wagner et al., 2012). The concept of mentalizing, or the ability to infer the mental states of others, has been consistently linked to heightened activity in a separate subsystem of the default mode network that includes the dmPFC and the right temporoparietal junction (TPJ) (Saxe, 2006; Schurz et al., 2014; Spreng et al., 2009; Van Overwalle, 2009).

This distinction seems to suggest a prominent line between brain processing for self and other, however, this line gets fuzzy when the identity of the other is experimentally manipulated. Mitchell et al. (2006) initially investigated this by controlling how similar or dissimilar the

particular stimuli representing others were to the self. They found that similar others tend to show heightened activity in vmPFC, the region usually implicated in self-referential cognition, whereas heightened dmPFC activity was found for the dissimilar others. This demonstrated that similarity could be driving self-other overlap within the vmPFC and suggested that the ability to understand similar others may require the use of self-knowledge. However, Krienen et al. (2010) provided further clarification by showing that it wasn't similarity, per se, that was driving this self-other overlap but was instead closeness to the self in terms of social relationships that was the biggest driver of heightened activity for others within the vmPFC.

Some have suggested that the engagement of the default mode network, and particularly regions involved in self-reference, for mentalizing or forming impressions of others inherently involves a form of self-projection or self-simulation. Tamir & Mitchell (2010) provided evidence for this claim by showing that a particular part of the mPFC, one that lies between the dorsal and ventral portions, shows a linear increase in activity as a function of the degree of similarity between one's own preferences and opinions and those inferred for a particular other. They suggested that we engage in anchoring and adjusting processes as we judge the similarity or differences that are present between a judged other and ourselves and showed that the mPFC is an important part of this process. Meyer (2020) provided evidence for this self-simulation account as well, such that the similarity of multivariate patterns of self were more similar to others as a function of social closeness. However, the overall RSA model, showed that patterns in the default mode network were significantly dissociated into three distinct categories of self, close others, and celebrities. This suggested that although there may be similarities between self and other at a representational level, the self is still a distinct construct and is likely not conflated at a neural level with representations of others.

While it is likely distinct from the identity model that the brain creates for the self, there is increasing evidence that the brain not only differentiates between close and distant other, but may also create distinct identity models of particular close individuals. It is interesting to note though that the patterns of activity that represent the self are similar to the patterns that others have of the perceiver when averaged across individuals in a distinct social network (Chavez & Wagner, 2020). Hassabis (2014) demonstrated that the dmPFC contained patterns of activity that could reliably be used to distinguish between four different fictional identities suggesting that there were patterns of activity that dissociated the identities of the four characters. Thornton and Mitchell (2017) extended these findings by using 20 distinct personally familiar others and had the participants imagine these people in various situational contexts. Using RSA, they showed that the representational similarity structure was best modeled by the identity of the individuals and not by the situations that they were embedded in. Relating this to prominent psychological theories of person perception, they found that the relational models theory (Fiske & Haslam, 1996), which describes the individuals in terms of the form of relationship they have with the participant, significantly fit the data. This suggested that the brain may be forming identity models of particular close others that represent the ways in which their relationship is important to the perceiver's sense of self and well-being.

These close relationships have also been shown to modulate the ways in which the brain engages with social information in general and in the formation process of the representational models of others. Parkinson et al. (2018) demonstrated the ways in which social network distance can affect neural synchrony during naturalistic video viewing. The study employed the use of media clips from a wide range of topics and used an intersubject correlation method to calculate the degree to which voxel time series patterns were consistent across subjects. The

strength of these correlations was predictive of the degree of distance in a social network such that people that had closer relationships with one another demonstrated higher levels of neural synchrony while watching the videos. Similarly, Guthrie et al. (2022) investigated the effects that social relationships had on the similarities between neural representations that two individuals may have of the same third individual. The study found that higher degrees of social closeness between the two individuals resulted in patterns of activity that were more similar while thinking about the third individual suggesting that social relationships may have a modulatory effect on the encoding of social information.

This work collectively shows that social relationships have an impact on the way that the brain processes information. It is likely that close relationships are prioritized in the brain as a form of self-relevance due to the importance of maintaining the relationship over time (Fareri et al., 2012; Krienen et al., 2010; Pfeifer & Berkman, 2018). To properly manage this social information the brain has developed a strategy for modeling the distinct attributes of particular others (Thornton & Mitchell, 2018), possibly in relation to the self (Tamir & Mitchell, 2010), for the purpose of predicting how that individual is going to act in the future and how their behavior is going to affect any long-term goals a perceiver may have (Tamir & Thornton, 2018). This suggests that close others may be used as an operationalized form of self-relevant processing to investigate the ways in which social contextual information is prioritized in a self-referential way.

Summary

The concept of the self has been a topic of inquiry for thousands of years, bringing with it rich philosophical and cognitive theoretical foundations. These introspective and behavioral

investigations revealed that the self must be both stable yet capable of change (Beauvoir et al., 1989; Locke, 2000; Markus, 1977; Sartre & Sartre, 2012), accessible through self-reflection (Aristotle, 1999; Descartes & Cottingham, 2013; Kuiper & Rogers, 1979; Plato & Segal, 1986), intimately connected to our social engagement (Fiske & Neuberg, 1990; Mead, 1934) and able to engage with and coordinate nearly all the important high level cognitive processing streams such as memory, attention, and cognitive control (Markus, 1977; Markus & Kitayama, 1991; Rogers et al., 1977). The past two decades of neuroscientific experimentation, with the help of neuroimaging technologies and computational modeling approaches, have provided grounded mechanistic support for many of these claims. The evidence points to the self as a neural construct that is unique and differentiated from purely semantic schemas (Craik et al., 1999; Kelley et al., 2002; Northoff et al., 2006; Wagner et al., 2012) and is powerful, in its own right, such that it has the ability to modulate the way in which sensory information is processed and prioritized (Berkman et al., 2017; Heatherton, 2011; Humphreys & Sui, 2016).

This dissertation will serve to build upon this foundation of knowledge by providing empirical evidence supporting the hypothesis that the self is a unique and powerful cognitive construct that facilitates the recruitment of various cognitive resources to deeply process self-relevant information and modify existing knowledge structures. The subsequent chapters will present a series of studies investigating the neural processing of self-relevant and socially relevant information and will explore topics such as the neural processing of self and others in naturalistic settings (Chapter 2), the influence of social relationship strength on the creation of unique or normative representations of others (Chapter 3), and the decoding of person identity for personally familiar others (Chapter 4). By integrating the findings from these studies, this dissertation seeks to demonstrate the central role of the self in shaping cognitive processing and

social understanding and will emphasize the need to consider the self not only in the context of social cognition but also in the context of general cognitive principles.

This dissertation includes coauthored material that has been previously published. Specifically, Chapter 3, titled *"Normativity vs Uniqueness: Effects of Social Relationship Strength on Neural Representations of Others,"* was co-authored with Robert S. Chavez. This work has been published in Social Cognitive and Affective Neuroscience. I was the lead author, responsible for writing the manuscript and conducting the methods and analyses, with input and revisions from my co-author.

Chapter II
Dyadic Neural Representations of Familiar Others

This work was co-authored with Jack Kapustka, Nate Holley & Robert S. Chavez. I was the lead author of the publication and responsible for the writing of the manuscript with edits provided by the co-authors. I led the study design, methods and analyses with input and assistance from the co-authors.

Introduction

Every moment of our waking lives, our brains are being bombarded with an overabundance of sensory information from our environment. This information, however, is not all equally important to us and the vast majority of it is filtered out before it ever gets to conscious awareness (Desimone & Duncan, 1995). The degree to which our brain attends to, processes, and remembers information is heavily influenced by the personal relevance of the information itself (Humphreys & Sui, 2016; Spreng & Grady, 2010). It is usually the stimuli that hold significant value or meaning to our lives, goals, and well-being, that get prioritized, deeply processed and made available to the mechanisms of awareness and self-reflection. (Beer et al., 2010; Heatherton, 2011; Meyer & Lieberman, 2018; Northoff & Hayes, 2011). The needs we have as humans are far more abstract than that of other animals and are strongly influenced by our social environment (Maslow, 1943; Pfeifer & Berkman, 2018). While some studies have demonstrated that the value we place on these close social relationships has a modulatory effect on brain processes for static controlled stimuli (Guthrie et al., 2022; Heatherton et al., 2006; Krienen et al., 2010), there is a lack of understanding, outside the use of preexisting media (Baldassano et al., 2018; Chen et al., 2017a; Hasson, 2004; Hasson et al., 2010), regarding how self-relevance can modulate the way we attend to and process naturally unfolding social information.

Research has consistently shown that people pay heightened attention to stimuli that are personally relevant or valuable to them in some way. For example, studies using the attentional

blink paradigm have shown that self-relevant stimuli, such as someone's own name, are more likely to be detected and processed (Mack & Rock, 1998; Shapiro et al., 1997). Similarly, research on the cocktail party effect has revealed that, when in noisy environments, people are more likely to notice and attend to their own name or personally relevant information even when they are engaged in another conversation (Cherry, 1953; Moray, 1959). More recent work has extended these findings with results that show that self-relevant stimuli, such as one's own face or name, capture attention more readily and are processed more efficiently than non-self-relevant stimuli (Alexopoulos et al., 2012; Sui et al., 2006). This suggests that stimuli that have self-relevant qualities can preferentially capture attention and are processed more deeply.

Neuroimaging studies have extended these behavioral findings and provided valuable insights into the attentional effects of self-relevance and their neural underpinnings. This self-reference advantage has been linked to increased activity in various brain regions involved in the perceptual features of the stimulus itself, including the fusiform face area (Sui et al., 2006), and regions involved in self-referential processing more generally such as the vmPFC (Heatherton et al., 2006; Kelley et al., 2002; Sui et al., 2013). Notably, Sui and Humphreys (2016) have proposed a Self-Attention Network model, which suggests that the vmPFC, along with other regions such as the posterior superior temporal sulcus and the intraparietal sulcus, form a network that supports rapid detection and preferential processing of information that has self-relevant qualities. This evidence points to a self that is integrated into multiple networks giving it the capability of utilizing attentional resources and facilitating the special status that self-relevant stimuli receive.

The modulatory effects of personal relevance extend beyond the self and into the social domain, particularly when considering close others. Thinking about close others, such as friends and family members, also elicits distinct cognitive patterns compared to thinking about unfamiliar others (Aron et al., 1991; Sui et al., 2012; Symons & Johnson, 1997). This suggests that the relevance and importance of close others and the relationships we maintain with them can also influence cognitive processes, but to a lesser extent than information that is purely self-relevant. Studies have shown that thinking about close others can also lead to preferential processing and attentional capture, with individuals exhibiting better memory for information related to close others compared to unfamiliar others (Bower & Gilligan, 1979; Symons & Johnson, 1997) and showing faster and more accurate identification of stimuli associated with close others (Keyes & Brady, 2010; Sui et al., 2012). This highlights the social aspects of self-processing in that the personal relevance of close others can enhance memory encoding and create an attentional bias towards this information in a similar way.

At the neural level, thinking about close others has been shown to engage brain regions that partially overlap with those involved in self-referential processing, such as the vmPFC, which is more active when making judgments about similar others compared to dissimilar others (Krienen et al., 2010; Mitchell, Macrae, et al., 2006). This suggests that the vmPFC is sensitive to the degree of self-relevance in social cognition and not distinctly self-specific information. Additionally, social relationships have been shown to modulate neural synchrony as well during shared experiences (De Felice et al., 2024; Golland et al., 2015; Parkinson et al., 2018). The degree of neural synchrony in brain regions involved in social cognition, such as the dmPFC and TPJ, can predict the degree of friendship between individuals, with closer friends exhibiting higher neural synchrony compared to more distant friends or strangers (Parkinson et al., 2018).

The use of stimuli with naturalistic qualities, such as movies, narratives, or spoken stories, has been on the rise in fMRI research to help investigate how the brain processes information in a more ecologically valid manner. The introduction of intersubject correlation (ISC) analysis has allowed for a new approach in studying the brain's response to these temporally dynamic stimuli by measuring the similarity of brain activity across individuals that are seeing or hearing the same stimulus (Hasson, 2004; Hasson et al., 2008; Lerner et al., 2011). This is done by calculating the correlation of the voxel time series in one individual's brain to the corresponding voxel time series in a second individual's brain iteratively across all relevant voxels. This allows for the identification of brain regions that respond consistently to the stimulus across participants and provides a new means of examining the temporal dynamics of neural processing.

The use of ISC analysis in fMRI research has provided valuable insights into how the brain processes narrative information (Baldassano et al., 2017; Chen et al., 2017b; Hasson et al., 2010; Lerner et al., 2011; Nastase et al., 2019). A key finding in this domain is the existence of temporal receptive windows, or the differing time scales over which different brain regions integrate information (Hasson et al., 2008; Lerner et al., 2011). This work has identified a hierarchy of processing steps, with lower-level sensory areas exhibiting shorter temporal receptive windows and higher-order brain regions, including key nodes of the default mode network, displaying longer temporal receptive windows (Baldassano et al., 2017; Hasson et al., 2015). These higher-order regions, which have been implicated in social cognition and self-referential processing, have been shown to be involved in the integration of information over long periods of time, allowing for the processing of more abstract and temporally distant features of the narrative (Hasson et al., 2015; Simony et al., 2016). This hierarchical organization enables

the brain to build increasingly rich and coherent representations of the unfolding story over time, as it integrates incoming sensory information with the broader knowledge of the narrative and related content from memory.

While the use of ISC analyses with naturalistic stimuli has provided valuable insights into the general principles of narrative processing in the brain, the majority of these studies have relied on preexisting media, such as movies/shows, audiobooks or radio programs, as stimuli (Ben-Yakov et al., 2012; Chen et al., 2017a; Hasson, 2004; Hasson et al., 2008; Honey et al., 2012; Lerner et al., 2011; Nastase et al., 2019). These stimuli fulfill the purpose of providing insight into general semantic or narrative processing, but they may not fully capture the effects that personal relevance and social significance may have on the ways in which the information is being processed. There is a need to extend this line of research by investigating how the brain responds to narratives that are directly relevant to the individual and their social relationships and connect these findings to the cognitive and neuroscience theories of self-related processing.

By employing a novel paradigm that uses personalized narratives from the participants themselves, their close others, and unfamiliar individuals, we can investigate the modulatory effects of personal relevance on the neural processing of naturalistic stimuli. The stranger condition, which is most similar to the stimuli used in previous literature, serves as a critical comparison point to examine how self-relevance and familiarity influence brain activity and intersubject correlation. Using univariate analysis methods, we hypothesize that self-relevant narratives (self or dyadic partner) will elicit heightened activity in areas involved in auditory sensory processing, consistent with attentional effects on low level sensory areas observed in other studies (Hopfinger et al., 2000; Sui et al., 2013). We also expect increased activation in

brain regions associated with social cognition and self-referential processing, including key nodes within the default mode network (Buckner et al., 2008; Heatherton, 2011; Northoff et al., 2006; Schmitz & Johnson, 2007; Wagner et al., 2012), when participants listen to personally relevant narratives compared to stories from unfamiliar individuals.

In regard to the ISC analyses, we predict that self-relevant narratives will result in higher levels of neural synchrony, compared to narratives from unfamiliar individuals. Specifically, we anticipate increased ISC in default mode network regions and attentional networks for self and partner conditions relative to the stranger condition. This hypothesis is supported by previous findings demonstrating enhanced neural synchrony between individuals who share personal experiences or have close social connections (Parkinson et al., 2018). By investigating this from both a univariate and an ISC approach, this study aims to provide a more comprehensive understanding of how the brain processes socially and personally relevant information and how it engages networks involved in attention and the hierarchical processing of naturalistic sensory information. The findings will help to bridge the gap between findings on general narrative processing and those regarding self-relevant prioritization.

Methods

Participants

A total of 58 adult subjects (29 dyads) were recruited for the study from the University of Oregon and the surrounding area. Each dyad consisted of two individuals who were at least familiar with one another (not strangers), with varying types and strengths of relationships (e.g., friends, romantic couples, work acquaintances). Participants ranged in age from 18 to 48 with a mean age of 23.8 and a standard deviation of 6.7. A majority of participants identified as White

(79.3%), 13.8% identified as Asian, and 3.4% chose not to identify their race. One participant identified as Black or African American. A majority of participants identified as non-Hispanic (86.2%). Participants' gender identity included 44.8% identifying as male (cisgender or transgender), 55.2% identifying as female (cisgender or transgender), and 1.7% identifying as gender queer. Regarding sex, 39.7% identified as male, and 60.3% identified as female.

All participants were screened for MRI contraindications and met the following inclusion criteria: age 18-60 years old, right-handed, proficient in English, normal or corrected-to-normal vision, no history of neurological disease, not pregnant, and no recreational drug use before the MRI scan. Participants with permanent dental retainer wires were excluded due to potential imaging artifacts. This study employed a three-part design, consisting of an online behavioral questionnaire, an interview session, and an MRI session. Each participant completed all three sessions individually. All participants provided informed consent in accordance with the guidelines set by the Institutional Review Board at the University of Oregon and were compensated for their participation.

Study Design

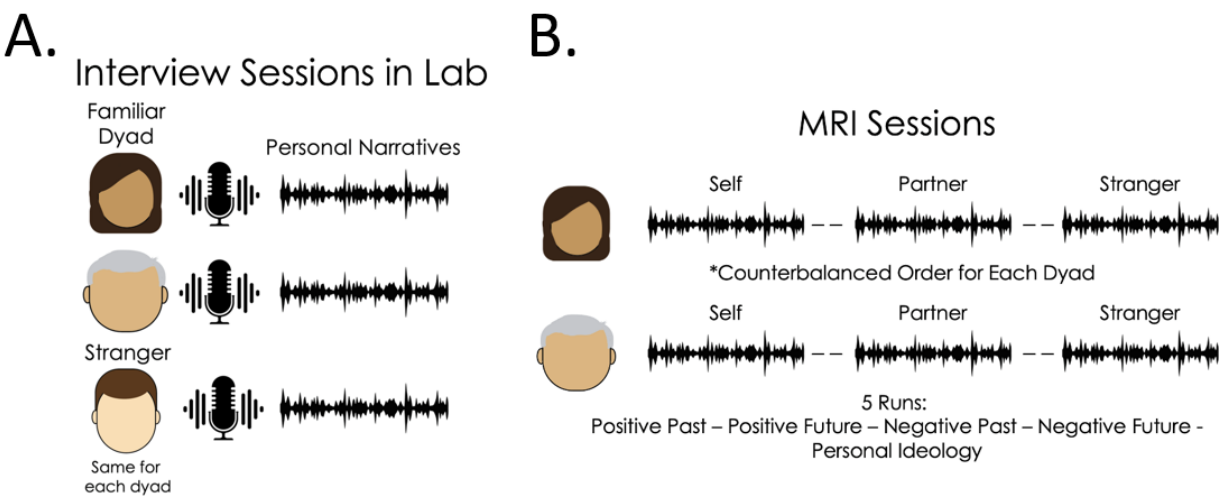

Figure 1. Experimental design overview. A) Interview sessions: Dyadic partners were interviewed individually in the lab, where they provided personal narratives on five prompts (positive past experience, potential positive future, negative past experience, potential negative future, and personal ideology). The same interview process was conducted with a recruited stranger participant, whose narratives were used as the stranger condition for all subjects. All narratives were recorded for use in the subsequent MRI portion. B) MRI sessions: Each subject was scanned independently, completing five runs corresponding to the five narrative prompts. In each run, the subject listened to their own narrative, their partner's narrative, and the stranger's narrative for that prompt. The order of self, partner, and stranger narratives was consistent within dyads but counterbalanced across dyads.

Online Behavioral Questionnaire

Participants first completed a self-paced online questionnaire using the Qualtrics platform. The questionnaire took approximately one hour to complete and included a consent form detailing the participants' rights, including the right to withdraw from the study at any time without penalty. The questionnaire consisted of two main sections: self-related questions and dyadic-partner related questions. In the self-related section, participants answered a series of questions about themselves using slider bars to indicate their agreement or disagreement with each statement. The dyadic-partner related section included questions about the participant's dyad partner and their relationship with that person, also using slider bars. Participants were instructed to find a quiet, distraction-free environment to complete the survey and to provide their best guess if unsure about the most appropriate response. Upon completion, participants

received a $10 Amazon gift card via email. The ratings collected in this session were not used in the present analysis.

Interview Session

Next, participants attended an individual, semi-structured interview with a trained researcher at the University of Oregon (Figure 1A). Interviews were conducted in a consistent and designated room and lasted approximately one hour. Participants were asked to arrive 10 minutes early to complete paperwork, including a consent form and a COVID-19 contact tracing form. The researcher explained the purpose of the interview, emphasizing that the sole purpose was to collect data concerning people's life stories.

The interview consisted of five main questions, adapted from the McAdams life story model (McAdams, 1996, 2001, 2018), each designed to elicit a detailed narrative response lasting approximately five minutes. Participants were encouraged to come prepared with ideas for their responses to maximize the allotted time and to provide a rich, detailed account of their experiences. The questions covered the following topics: (1) a peak experience (high point) from the participant's past, (2) a nadir experience (low point) from their past, (3) a positive future event, (4) a negative future event, and (5) the participant's personal ideology and values. For each question, the researcher used prompts to encourage participants to elaborate on their thoughts, feelings, and the impact of the events they described.

Interviews were audio-recorded using a high-fidelity microphone, and participants were informed that their responses would be heard by their dyad partner and the research team. The researcher maintained a neutral demeanor throughout the interview, using non-verbal cues to show engagement and understanding while minimizing verbal responses that could influence the participant's narrative. The researcher used a timer to track the length of responses and raised

their hand at the 5-minute mark to signal the participant to conclude their thought. Upon completion of the interview, participants were compensated with $10 cash and scheduled for their MRI session. The audio recordings of the interviews were then prepared for use in the subsequent MRI session**.**

Audio Stimuli Preparation

The audio recordings from the interview sessions were processed using Adobe Audition 2022 (Adobe Inc., 2022) to create the stimuli for the MRI session. Five separate audio files for each prompt (Positive Experience, Positive Future, Negative Experience, Negative Future, and Personal Ideology) were created for each subject. First basic noise reduction was applied by capturing a noise print from a short clip without any speech and applying it to the entire file. Next, any prompts or interruptions from the interviewer were removed from the audio file, leaving only the participant's responses. The volume was then adjusted to match the volume of a pre-selected "Stranger Condition" recording, which served as a baseline. This step ensured that all audio clips were at a consistent volume level during the MRI session.

MRI Session

Lastly, participants completed an individual fMRI session at the Lewis Center for Neuroimaging at the University of Oregon. The session lasted approximately two hours, with about 90 minutes spent inside the scanner. Participants were asked to arrive 10 minutes early to complete paperwork, including an MRI safety screening form and a consent form. The researcher reviewed the screening form with the participant to ensure their safety and compatibility with the MRI environment.

During the MRI session, participants listened to audio recordings of their own personal narratives, their dyad partner's narratives, and a stranger's narratives while functional images of

their brain were acquired (Figure 1B). The task consisted of five runs, each containing three approximately 5-minute stories with 15 seconds of silent fixation between them. The three stories heard in each run all corresponded to the same prompt from the interview session. For example, all three stories in the first run were positive past stories from the participant, their partner, and a stranger. The stranger narratives were consistent across all subjects in the study and were created using the same interview prompts with a participant recruited specifically for this purpose. Participants were instructed to keep their eyes open and fixed on a cross displayed on a screen during the task to ensure engagement with the task. The order in which participants heard the stories (self, partner, or stranger) was randomized across dyads to minimize potential order effects, but members of the same dyad heard the stories in the same conditional order.

Upon completion of the MRI session, participants then completed an exit interview, which involved answering questions about the content of the audio narratives they heard in the scanner, rating their emotional experiences, and indicating any personal connections to the stories. These measures were not used in the present analysis. Finally, participants were compensated with $40 cash.

MRI Data Acquisition

MRI was conducted with a Siemens 3T Prisma scanner using a 64-channel phased array coil. Structural images were acquired using a T1-weighted MP-RAGE protocol (176 sagittal slices; time repetition [TR]: 2500 ms; time echo [TE]: 3.43 ms; flip angle: 7°; 1-mm isotropic voxels). Functional images were acquired using a $T2^*$-weighted echo planar sequence (TR: 1500 ms; 72 axial slices; TE: 25 ms; flip angle: 90°; 2-mm isotropic voxels). For each participant, we collected five runs of the narrative task (Approximately 630 whole-brain volumes per run – narratives were slightly variable in length). In order to correct for distortion due to B0

inhomogeneity, we also acquired a field map (TR: 4690 ms; TE: 33.2 ms; effective echo spacing: 0.265 ms). The total length of time for the entire scanning session was approximately 90 minutes, and each of the five functional runs was approximately 16 minutes long.

Neuroimaging Analysis

Preprocessing

Functional MRI data underwent a series of preprocessing steps using FSL (S. M. Smith et al., 2004). First, the data were subjected to a mean-based intensity normalization and high-pass filtering using a Gaussian-weighted least-squares straight line fitting with a sigma of 140 seconds. The data were then spatially smoothed using a 6-mm full-width at half-maximum (FWHM) Gaussian smoothing kernel.

To register the results to standard space, a multistep normalization procedure was employed. Initially, functional data were corrected for spatial distortion using a field map unwarping technique. The corrected functional data were then aligned to each participant's anatomical scan using boundary-based registration (Greve & Fischl, 2009) in conjunction with a linear registration using FSL's FLIRT (FMRIB's Linear Image Registration Tool). The resulting images were warped to a 2-mm Montreal Neurological Institute (MNI) template using nonlinear registration with FSL's FNIRT (FMRIB's Nonlinear Image Registration Tool) and a 10-mm warp field. All subsequent preprocessing steps were performed in native space before being warped into standard space for group-level analyses.

To prepare the data for intersubject correlation analyses, the functional files in native space were divided into three separate nifti files, each corresponding to one of the three audio stories plus 15 seconds of fixation at the end. The exact start time of the audio was determined for each of these files, and slice-time correction was used to interpolate the data, aligning it with

the beginning of the audio. This was done by subtracting the start time of the audio (e.g., 0.56 seconds) from all values in the slice-time file, effectively resetting the 0 mark and ensuring that the audio start time was treated as the beginning of the repetition time (TR).

This slice-time correction step was crucial to ensure that all nifti files were temporally aligned across participants within a dyad. For example, if subject 1 in a dyad had their stranger audio start at 0 seconds, while subject 2 in the same dyad had the same audio start at 0.56 seconds, the adjustment was necessary to guarantee that the time series information was properly aligned for subsequent intersubject correlation analyses. After the files were cut and slice-time corrected, they were then transformed into MNI standard space. The transformation parameters used for this step were derived from the registration of each participant's functional data to their anatomical data and then to the standard space during the preprocessing stage. This ensured that all participants' data were in the same standard space, allowing for group-level analyses and comparisons.

Univariate Analysis

To investigate differences in brain activation across the three conditions (self, partner, and stranger), a univariate analysis was conducted using a block design approach. The audio stimuli for each condition in each run were modeled as blocks with a duration of approximately 5 minutes, allowing for the examination of sustained activity that was significantly different between conditions. Each run consisted of three blocks, one for each condition (self, partner, and stranger).

The analysis was performed using FMRIB Software Library (FSL) version 6.0.1 (Jenkinson et al., 2012). At the first level, general linear models (GLMs) were constructed for

each participant, with the three conditions modeled as explanatory variables (EVs) and convolved with a double-gamma hemodynamic response function. The contrasts of interest were defined as follows: (1) all pairwise combinations of the conditions (self > partner, self > stranger, partner > self, etc.), and (2) dyadic information versus stranger information (self + partner > stranger and stranger > self + partner, with self and partner equally weighted). The latter set of contrasts was designed to identify brain regions where the processing of information specific to the dyad differed from that of stranger-related information, which was particularly relevant for the subsequent intersubject correlation analysis.

At the second level, a fixed-effects analysis was performed to combine the parameter estimates across the five runs (positive past, positive future, negative past, negative future, and personal ideology) for each participant. This step yielded a single set of contrast images per participant, representing the overall effect of each contrast across the different story types. Finally, a group-level analysis was conducted using a mixed-effects model (FLAME 1) with cluster correction to estimate the strength of the effects across all subjects. The cluster-defining threshold was set at $Z > 3.1$, and the cluster significance threshold was set at $p < 0.05$, corrected for multiple comparisons using Gaussian random field theory. This approach accounted for both within-subject and between-subject variability while controlling for multiple comparisons.

Intersubject Correlation Analysis

An ISC approach was employed to identify brain regions exhibiting synchronized activity between dyad members when listening to the same audio stimuli (Hasson, 2004; Nastase et al., 2019). For each dyad, the preprocessed and spatially normalized functional data were paired according to the audio condition (Figure 2A). Specifically, for each run, the self-condition of

subject 1 was paired with the partner-condition of subject 2, the partner-condition of subject 1 was paired with the self-condition of subject 2, and the stranger-conditions of both subjects were paired together.

ISC analysis was performed using custom scripts written in Python. For each voxel, the Pearson correlation coefficient was computed between the time series of the paired subjects, resulting in a 3D correlation map where each voxel value represented the strength of the temporal synchronization between the two subjects at that spatial location (Figure 2B) (Nastase et al., 2019). This process was repeated for each dyad and each run, yielding three ISC maps per run: self-partner, partner-self, and stranger.

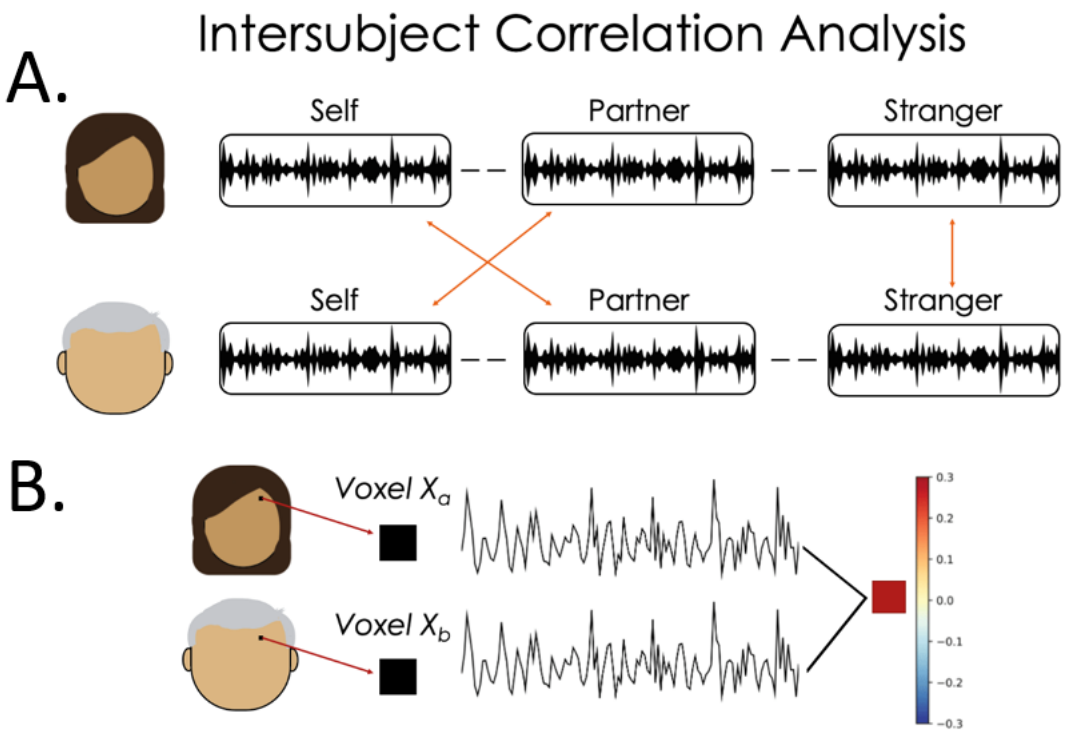

Figure 2. Intersubject correlation (ISC) analysis methodology. A) Functional MRI data alignment: To enable ISC analysis, the brain activity time series were temporally aligned across dyadic partners based on the matching audio narratives. For example, subject 1's "self" condition corresponds to the same audio file as subject 2's "partner" condition. B) ISC computation: For each voxel in standard brain space, the time series of neural activity is extracted for both subjects in a dyad and correlated, yielding an ISC value representing the degree of neural synchrony between the subjects while listening to the same narrative. This process is repeated for all voxels, resulting in a 3D map of ISC values throughout the brain. For each experimental run, three ISC maps were computed: self (subject 1) / partner (subject 2), partner (subject 1) / self (subject 2), and stranger.

Parcellation and Group Level ISC Analysis

To further investigate the differences between conditions across brain regions, the ISC output for each condition and each run (3 outputs for each of the 5 runs) was parcellated using the Schaefer 400 parcel scheme (Schaefer et al., 2018). This parcellation scheme divides the brain into 400 functionally distinct cortical regions, allowing for a more focused examination of regional differences in intersubject correlation strength. The average ISC correlation was then calculated within each of these regions, and this average correlation was used as the dependent measure for a linear mixed effects modeling approach.

To test whether there was a significant difference between self-relevant conditions (self/partner and partner/self) and the stranger condition in each of the 400 parcels, separate linear mixed effects models were run in each of the 400 parcels. The primary hypothesis of the study was that self-relevant conditions (self/partner and partner/self) would elicit significantly higher intersubject correlations compared to the stranger condition. The experimental design inherently produced two separate conditions (self/partner and partner/self) within the overarching self-relevancy construct. To effectively test the hypothesis and account for this design feature, the self/partner and partner/self conditions were collectively treated as a single self-relevant condition in the linear mixed effects modeling approach.

The data were dummy coded such that self/partner and partner/self conditions were both coded as 1, and stranger conditions were coded as 0. A linear mixed effects random intercepts model was then constructed for each parcel, with the average ISC in the given parcel as the dependent measure and the dummy coded condition column as the predictor. Since there were 5 runs per dyad that were not independent of one another, run nested within dyad was modeled as a random effect. This approach allowed for the consideration of the hierarchical structure of the data, accounting for the non-independence of observations within each dyad and run. If the model was significant and the beta weight was positive for the predictor, it would indicate that correlations were significantly higher for self-relevant conditions in that specific brain region compared to stranger conditions.

Results

Univariate Results

The univariate analysis revealed significant differences in brain activation across the three conditions (self, partner, and stranger). The most prominent finding from these analyses was the extensive activation observed when contrasting the self condition with both the partner and stranger conditions (Figure 3). Listening to one's own narrative elicited significantly greater activity across large portions of the brain, including key regions associated with the default mode network, language and auditory processing, and attention/salience networks. Heightened activity was also found in primary sensory regions, such as primary auditory cortex (A1) and the thalamus, while listening to self-narratives. This suggests that processing self-relevant information recruits a widespread network of brain regions, extending beyond the classic default mode network and encompassing sensory and other cognitive processing streams.

In contrast to the self condition, the partner condition yielded more limited activation differences when compared to the self and stranger conditions (Figure 3). Greater activity for the partner condition over the self condition was observed in a small region of the orbitofrontal cortex (OFC) and the right middle temporal gyrus. When comparing the partner condition to the stranger condition, greater activity was found in the ventromedial prefrontal cortex (vmPFC) and primary auditory cortex. This suggests that processing information related to a close other may also engage brain regions involved in social cognition and sensory processing, but to a lesser extent than self-referential processing.

The stranger condition showed greater activation than the self condition in regions similar to those observed in the partner > self contrast, including the OFC and middle temporal gyrus (Figure 3). Additionally, the temporal pole, amygdala and dorsal regions of the precuneus exhibited greater activity for the stranger condition compared to the self condition. A substantial number of regions showed greater activity for the stranger condition compared to the partner condition, including key regions of the limbic network such as the subgenual ACC, amygdala and the temporal pole, as well as various regions from the ventral attention/salience network. It is noteworthy though that the regions that show heightened activity for stranger vs self or partner all fall outside of the default mode network.

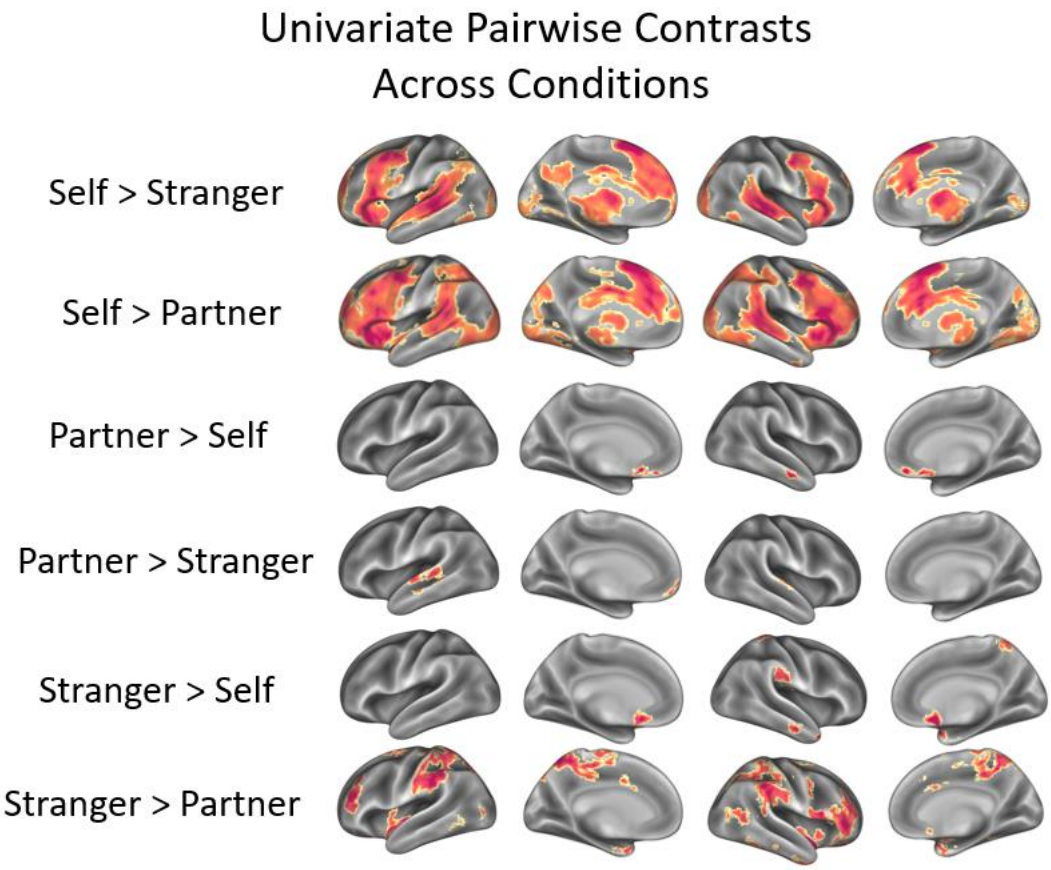

Figure 3. Univariate pairwise contrasts across conditions. Brain regions exhibiting significant differences in activity across all pairwise comparisons of the self, partner, and stranger conditions. Each condition represents the average neural activity while participants listened to narratives produced by themselves (self), their dyadic partner (partner), or an unfamiliar individual (stranger). The contrasts reveal differential engagement of brain areas associated with self-referential processing, social cognition, and the processing of novel information.

To further investigate the neural processing of self-relevant information, we examined the contrasts between the combined self and partner conditions (dyad) equally weighted in the model and the stranger condition. These contrasts revealed striking differences in the engagement of large-scale brain networks. The dyad > stranger contrast (Figure 4, Table 1, Figure 5) revealed widespread heightened activity across the default mode network, including key regions such as the medial prefrontal cortex, temporoparietal junction, precuneus, and hippocampus suggesting that these areas are more engaged when processing information related to oneself or a close other compared to a stranger. There was also heightened activity in areas involved in language and

narrative processing such as the superior temporal gyrus (STG)/primary auditory cortex (A1), the inferior frontal gyrus (IFG) and other areas along the superior temporal sulcus (STS). In stark contrast, the stranger > dyad contrast (Figure 4, Table 2, Figure 5) revealed a markedly different pattern of activation. The default mode network was notably absent, and instead, heightened activity was observed in the limbic network, particularly in the amygdala and subgenual ACC.

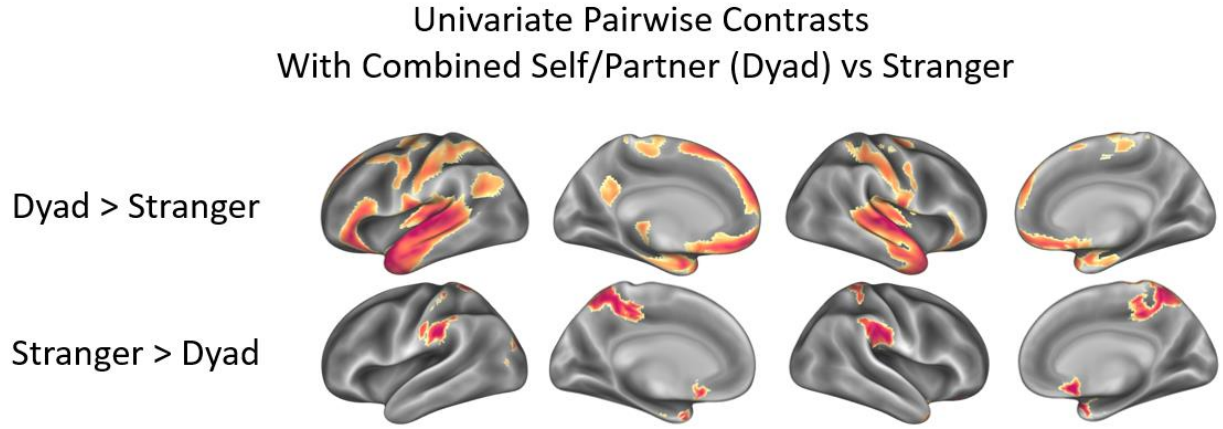

Figure 4. Univariate contrasts of combined self/partner (dyad) vs. stranger conditions. Brain regions showing significant differences in activity when contrasting the combined self and partner conditions (equally weighted) against the stranger condition. The dyad > stranger contrast reveals areas with greater activity for dyadic or self-relevant information, while the stranger > dyad contrast shows regions that were more active when processing information from an unfamiliar individual.

Table 1

Peak Activation Values in Significant Clusters Where Dyad (Self + Partner) Activation Was Significantly Greater Than Stranger

MNI Coordinates	Network	Hemisphere	Region	Z-stat
25, -80, -38	Cerebellum	Right	Cerebellum	9.1
8, -54, -43	Cerebellum	Right	Cerebellum	8.02
-23, -62, -55	Cerebellum	Left	Cerebellum	5.95

MNI Coordinates	Network	Hemisphere	Region	Z-stat
-60, -18, 0	Default Mode Network	Left	Superior Temporal Gyrus (STG)/(A1)	11.09
-47, 32, -11	Default Mode Network	Left	Inferior Frontal Gyrus (IFG)	9.67
-3, 55, -14	Default Mode Network	Left	Ventromedial Prefrontal Cortex (vmPFC)	9.3
59, -24, 2	Default Mode Network	Right	Superior Temporal Gyrus (STG)/(A1)*	9.28
-8, 36, 56	Default Mode Network	Left	Dorsomedial Prefrontal Cortex (dmPFC)	9.25
-3, 40, -19	Default Mode Network	Left	Orbitofrontal Cortex (OFC)*	8.91
-8, 27, 62	Default Mode Network	Left	Dorsomedial Prefrontal Cortex (dmPFC)*	8.02
3, 10, -14	Default Mode Network	Right	Orbitofrontal Cortex (OFC)	7.03
-22, -14, -16	Default Mode Network	Left	Hippocampus	6.52
-44, -58, 28	Default Mode Network	Left	Temporoparietal Junction (TPJ)	6.4
7, 56, 33	Default Mode Network	Right	Dorsomedial Prefrontal Cortex (dmPFC)	6.1
26, -13, -16	Default Mode Network	Right	Hippocampus	6.02
7, 55, -16	Default Mode Network	Right	Orbitofrontal Cortex (OFC)	5.71
40, 31, -16	Default Mode Network	Right	Inferior Frontal Gyrus (IFG)*	5.4
-6, -53, 26	Default Mode Network	Left	Posterior Cingulate Cortex (PCC)	4.94
35, -32, 44	Dorsal Attention Network	Right	Superior Parietal Lobule	5.74

MNI Coordinates	Network	Hemisphere	Region	Z-stat
-52, -26, 38	Dorsal Attention Network	Left	Superior Parietal Lobule*	5.67
-46, 13, -28	Limbic Network	Left	Temporal Pole	10.31
-64, -33, 12	Salience/Ventral Attention Network	Left	Posterior Superior Temporal Gyrus*	8.46
-41, -18, 6	Salience/Ventral Attention Network	Left	Middle Insula*	8.65
44, -12, 6	Salience/Ventral Attention Network	Right	Middle Insula*	8.36
-42, -16, 8	Salience/Ventral Attention Network	Left	Frontal Operculum*	8.14
-65, -49, 24	Salience/Ventral Attention Network	Left	Supramarginal Gyrus (SMG)*	3.26
-43, -22, 9	Somatomotor Network	Left	Middle Insula	10.88
-52, -18, 8	Somatomotor Network	Left	Temporal Operculum	10.74
49, -16, 6	Somatomotor Network	Right	Middle Insula	10.71
-23, -13, 57	Somatomotor Network	Left	Paracentral Lobule	5.75
-6, -26, 57	Somatomotor Network	Left	Paracentral Lobule	5.07
35, -27, 51	Somatomotor Network	Right	Postcentral Gyrus*	3.62

Note. This table presents the results from the univariate analysis where the contrast was set to self/partner (equally weighted) greater than the stranger condition. The coordinates correspond to peak activation values in significant clusters where dyad (self + partner) activation exceeded that of the stranger condition. Z-stat values are reported following corrections for multiple comparisons. The identified regions reflect the areas in which the peak activity values reside.
*Regions of overlap displayed in Table 4

Table 2

Peak Activation Values in Significant Clusters Where Stranger Activation Was Significantly Greater Than Dyad (Self + Partner)

MNI Coordinates	Network	Hemisphere	Region	Z-stat
60, -37, 42	Control Network	Right	Parietal Cortex	3.64
-55, -38, 48	Control Network	Left	Posterior Parietal Cortex	3.46
-20, -10, -21	Default Mode Network	Left	Hippocampus	4.06
4, -47, 61	Dorsal Attention Network	Right	Superior Parietal Lobule	5.23
-44, -41, 62	Dorsal Attention Network	Left	Superior Parietal Lobule	4.27
5, 20, -12	Limbic Network	Right	Subgenual ACC	5.66
20, -6, -17	Limbic Network	Right	Amygdala	4.95
-24, 1, -24	Limbic Network	Left	Temporal Pole	4.69
18, 36, -19	Limbic Network	Right	Orbitofrontal Cortex (OFC)	4.54
-6, 20, -13	Limbic Network	Left	Subgenual ACC	4.45
-17, 5, -21	Limbic Network	Left	Amygdala	4.23
22, 8, -31	Limbic Network	Right	Temporal Pole	4.05
-12, -50, 57	Salience/Ventral Attention Network	Left	Dorsal Precuneus	5.09
58, -28, 33	Salience/Ventral Attention Network	Right	Temporal Occipital Parietal Junction	4.88
-63, -31, 31	Salience/Ventral Attention Network	Left	Inferior Parietal Lobule	4.86
12, -35, 45	Salience/Ventral Attention Network	Right	Parietal Operculum	4.6

MNI Coordinates	Network	Hemisphere	Region	Z-stat
40, 8, -24	Salience/Ventral Attention Network	Right	Frontal Operculum	4.02
-57, -17, 34	Salience/Ventral Attention Network	Left	Inferior Parietal Lobule	3.53
26, -42, 62	Somatomotor Network	Right	Postcentral Gyrus	4.15
-22, -50, 65	Somatomotor Network	Left	Precuneus	4.06
-25, -47, 66	Somatomotor Network	Left	Paracentral Lobule	4.01
3, -26, 75	Somatomotor Network	Right	Postcentral Gyrus	3.78
-19, -36, 73	Somatomotor Network	Left	Paracentral Lobule	3.69
-53, -79, 11	Visual Network	Left	Lateral Occipital Cortex	4.15

Note. This table presents the results from the univariate analysis where the contrast was set to the stranger condition greater than self/partner (equally weighted). The coordinates correspond to peak activation values in significant clusters where the stranger condition activation exceeded that of dyad (self + partner). Z-stat values are reported following corrections for multiple comparisons. The identified regions reflect the areas in which the peak activity values reside.

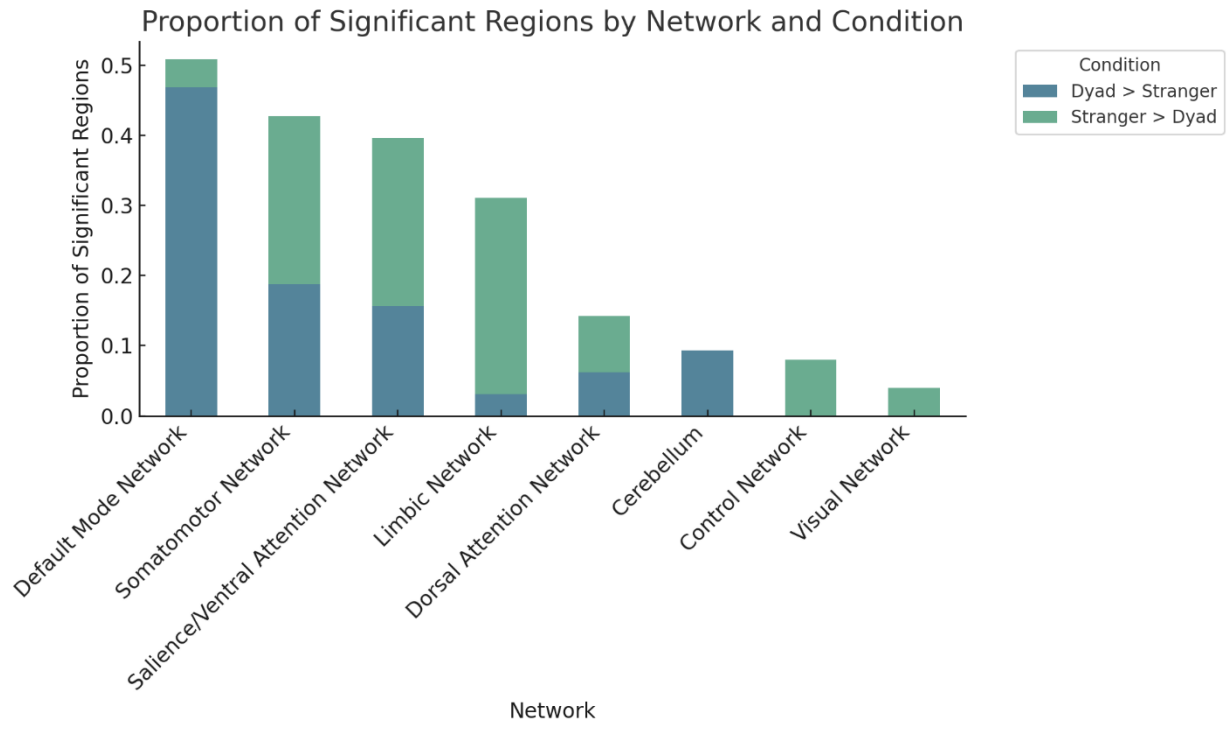

Figure 5: Count of significant regions by brain network for dyad > stranger and stranger > dyad contrasts. The bars represent the number of regions within each network that showed significant activation in either the dyad > stranger (blue) or stranger > dyad (green) contrast. Networks are ordered by total count of significant regions.

Intersubject Correlation Results

Intersubject correlation (ISC) analysis was employed to investigate neural synchrony between dyad members while they listened to each other's narratives and compare it to the synchrony found while dyad members were both listening to the stranger narratives. ISC maps were calculated for each corresponding audio file (Nastase et al., 2019), resulting in three ISC maps per run (self-partner, partner-self, and stranger) that were then parcellated into 400 distinct regions using the Schaefer 400 parcellation atlas (Schaefer et al., 2018). The average ISC values in each region were then used as the dependent measure in a series of mixed effects linear

models to examine differences in neural synchrony between dyad members and strangers. This approach allowed us to assess the differences in temporal synchronization in brain activity across participants for each of the conditions while accounting for the hierarchical structure of the data.

We first conducted the ISC analysis independently for each of the five runs, examining the differences in neural synchrony between the dyadic conditions (self/partner and partner/self) and the stranger condition. At this level of analysis, no regions survived multiple comparison corrections for significance. However, the raw beta weight estimates were plotted to visualize the magnitude and direction of the effects in each run (Figure 6). These plots provide valuable insights into the consistency and variability of the effects across different narrative contexts in different regions of the brain. This variability suggests that the strength and direction of neural synchrony between dyad members and strangers may depend on the specific content of the audio stimuli. However, the following combined analysis, which leveraged the increased power of the full dataset.

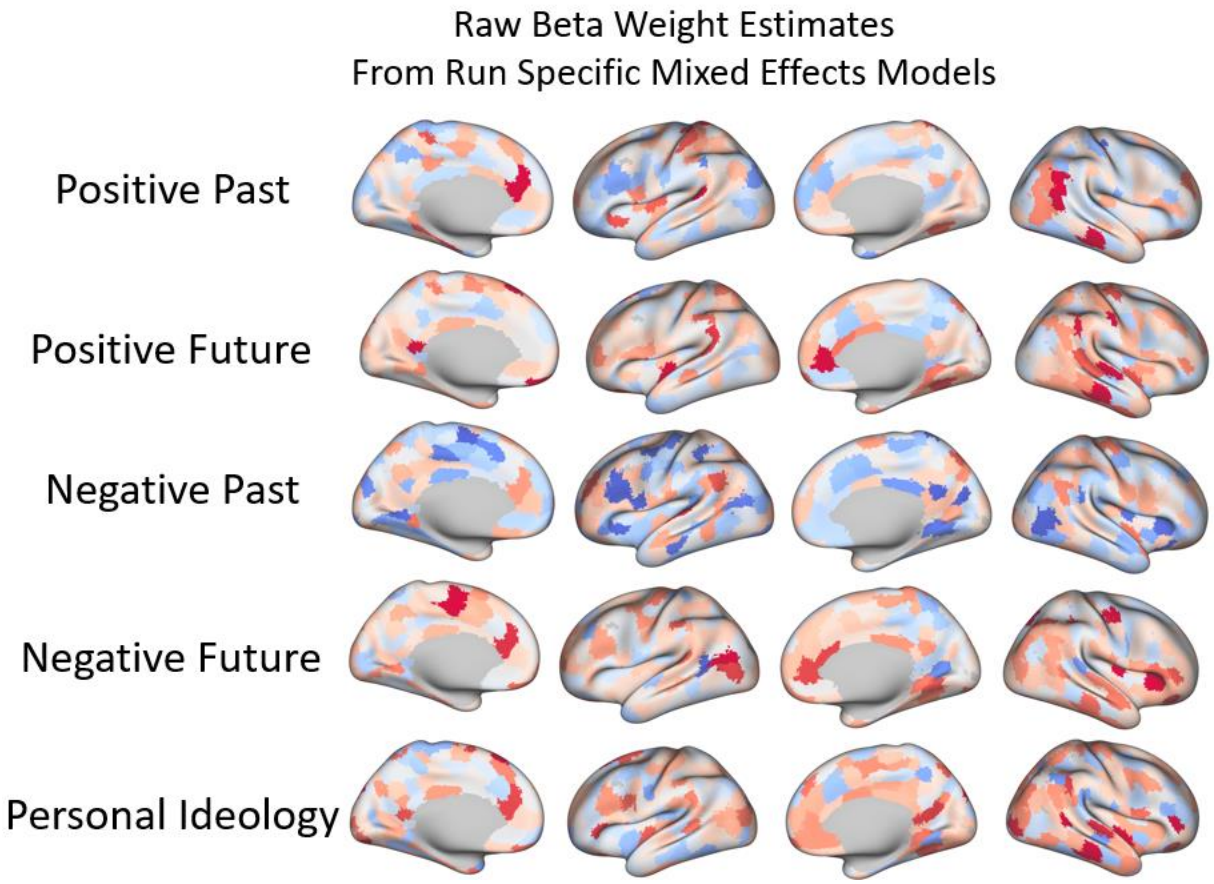

Figure 6. Raw beta weight estimates from run-specific mixed effects models investigating differences in intersubject correlation (ISC) values between dyadic (self/partner) and stranger narratives. Each row represents a different run, with the narrative prompts being positive past, positive future, negative past, negative future, and personal ideology. The brain maps depict the raw beta weights from the mixed effects models, with red indicating regions where ISC was greater for dyadic narratives compared to the stranger condition, and blue indicating the opposite effect. It is important to note that none of these findings reached statistical significance, and this figure is presented for illustrative purposes to demonstrate the variability across runs and the regions that showed consistent effects. The figure highlights the consistency of the effects across different story types and the potential regions of interest for further analyses.

To increase statistical power and enable the detection of more robust differences in neural synchrony, we constructed a second mixed effects model that combined data from all five runs. This analysis yielded a set of regions that survived multiple comparison correction, all of which exhibited greater neural synchrony in the dyadic conditions compared to the stranger condition

(Figure 7). Notably, no regions showed significantly higher ISC values for the stranger condition compared to the dyadic conditions. It is evident in the earlier run specific models that there were regions in which ISC was higher for the stranger condition than for the dyadic conditions but none of these effects survived multiple comparison correction at this level of the analysis.

The regions exhibiting enhanced neural synchrony for dyadic narratives compared to strangers were primarily located in brain networks associated with social cognition and attentional processes (Figure 7, Table 3, Figure 8). One of the most prominent findings was the enhanced ISC between dyad members in various regions within the default mode network including portions of the mPFC, PCC and TPJ. Key attentional/salience areas included portions of the insula and various areas of the operculum. Furthermore, many of the regions showing higher ISC for dyads are also implicated in language processing, such as the superior temporal gyrus/primary auditory cortex, inferior frontal gyrus, superior frontal gyrus, middle temporal gyrus, and supramarginal gyrus.

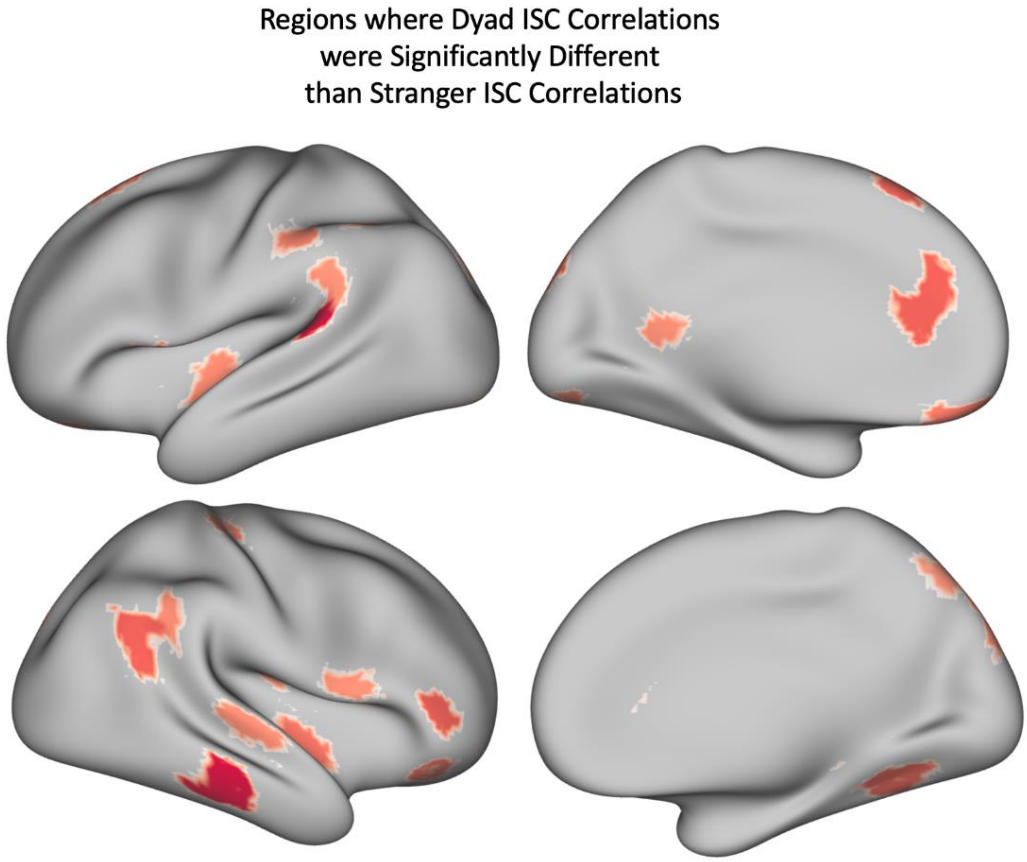

Figure 7. Regions where dyad ISC correlations were significantly different than stranger ISC correlations across all runs combined. The brain maps show areas where intersubject correlation values were consistently higher when participants listened to narratives from the self or partner (dyad) compared to narratives from a stranger. These results were obtained from a mixed effects model that combined data across all runs, increasing statistical power. The regions highlighted in red indicate where the dyad ISC values were significantly greater than the stranger ISC values after correcting for multiple comparisons. No regions survived multiple comparison correction in which the stranger ISC values were significantly greater than the dyad ISC values.

Table 3

Regions Showing Significant Differences in Intersubject Correlation (ISC) Between Self/Partner (Dyad) and Stranger Conditions

MNI Coordinates	Network	Hemisphere	Region	FDR Corrected p-value	Standard Error	Beta Estimate
48, 40, 4	Control Network	Right	Inferior Frontal Gyrus (IFG)	0.009	0.008	0.029
-42, -52, 48	Control Network	Left	Inferior Parietal Lobule (IPL)	0.03	0.005	0.017

MNI Coordinates	Network	Hemisphere	Region	FDR Corrected p-value	Standard Error	Beta Estimate
63, -27, -20	Default Mode Network	Right	Middle Temporal Gyrus (MTG)	0.005	0.009	0.035
-6, 34, 22	Default Mode Network	Left	Anterior Cingulate Cortex (ACC)	0.005	0.008	0.03
-8, -50, 9	Default Mode Network	Left	Posterior Cingulate Cortex (PCC)	0.005	0.006	0.022
-11, 27, 61	Default Mode Network	Left	Dorsomedial Prefrontal Cortex (dmPFC)*	0.009	0.008	0.03
55, -53, 26	Default Mode Network	Right	Temporoparietal Junction (TPJ)	0.009	0.008	0.029
35, 38, -12	Default Mode Network	Right	Inferior Frontal Gyrus (IFG)*	0.009	0.007	0.026
-9, 46, -23	Default Mode Network	Left	Orbital Frontal Cortex (OFC)*	0.015	0.008	0.027
53, -45, 34	Default Mode Network	Right	Temporoparietal Junction (TPJ)	0.037	0.007	0.022
63, -19, -1	Default Mode Network	Right	Superior Temporal Gyrus (STG)/(A1)*	0.039	0.007	0.021
-56, -31, 44	Dorsal Attention Network	Left	Superior Parietal Lobule (SPL)*	0.012	0.007	0.025
9, -75, 53	Dorsal Attention Network	Right	Precuneus	0.016	0.006	0.02
-59, -37, 17	Salience/Ventral Attention Network	Left	Posterior Superior Temporal Gyrus*	0.004	0.009	0.037
-44, 11, 1	Salience/Ventral Attention Network	Left	Frontal Operculum*	0.004	0.008	0.034
-38, -14, -1	Salience/Ventral Attention Network	Left	Middle Insula*	0.005	0.006	0.023
41, -11, -4	Salience/Ventral Attention Network	Right	Middle Insula*	0.009	0.007	0.026

MNI Coordinates	Network	Hemisphere	Region	FDR Corrected p-value	Standard Error	Beta Estimate
-59, -46, 26	Salience/Ventral Attention Network	Left	Supramarginal Gyrus (SMG)*	0.018	0.007	0.022
54, 11, 11	Salience/Ventral Attention Network	Right	Frontal Operculum	0.027	0.007	0.021
41, -13, 20	Salience/Ventral Attention Network	Right	Parietal Operculum	0.04	0.007	0.021
33, -27, 56	Somatomotor Network	Right	Postcentral Gyrus*	0.018	0.007	0.023
27, -50, -9	Visual Network	Right	Medial Temporo-Occipital Gyrus	0.004	0.007	0.028
-18, -86, -15	Visual Network	Left	Primary Visual Cortex (V1)	0.009	0.007	0.024
-17, -90, 35	Visual Network	Left	Superior Occipital Gyrus	0.012	0.008	0.03
18, -88, 37	Visual Network	Right	Superior Occipital Gyrus	0.05	0.008	0.023

Note. This table presents the regions where there were significant differences in inter-subject correlation (ISC) between self/partner (dyad) conditions and stranger conditions. The results are from a linear mixed effects model, and all regions listed here passed FDR correction for multiple comparisons. Positive beta estimate values indicate that the ISC correlations were higher for self/partner conditions compared to stranger conditions. The significance threshold was set at $p < 0.05$. *Regions of overlap displayed in Table 4

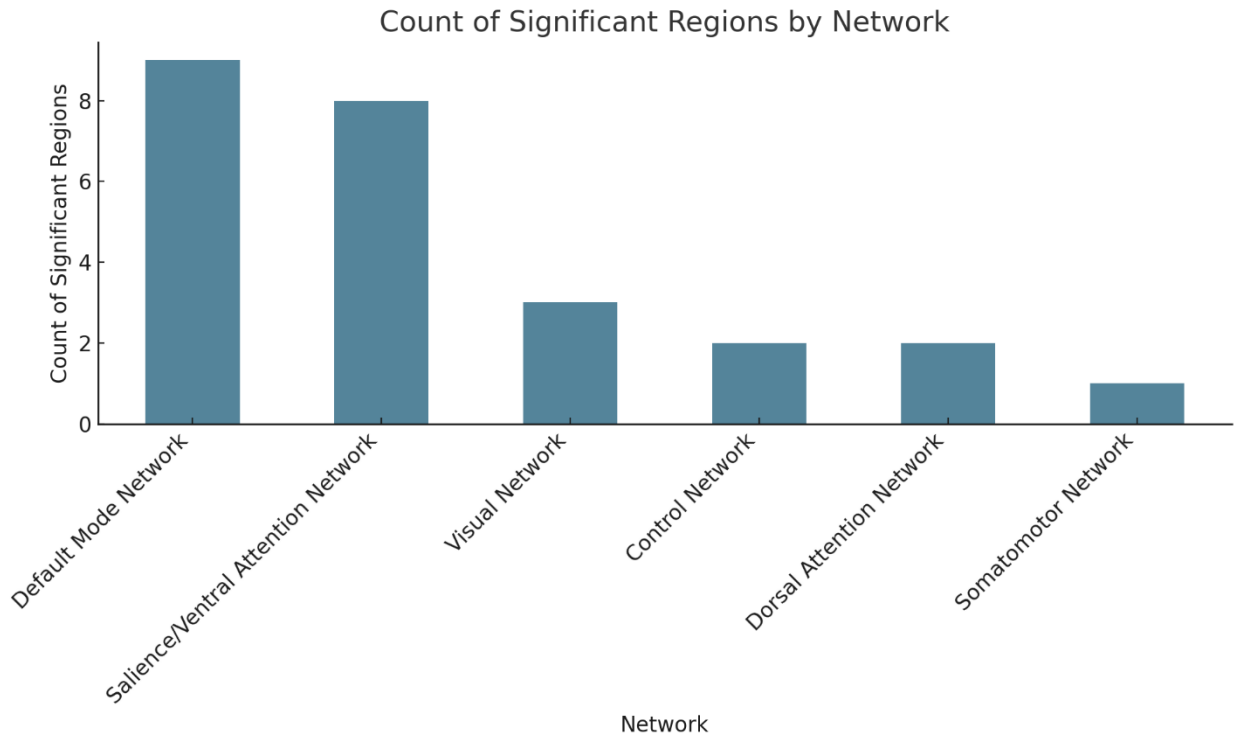

Figure 8: Count of significant regions by brain network where dyad ISC was significantly greater than stranger ISC. The bars represent the number of regions within each network that showed significantly higher ISC for dyad conditions compared to stranger conditions. Networks are ordered by the count of significant regions.

To investigate convergent evidence of neural mechanisms underlying the processing of self-relevant information in dyads, we examined the overlap between the univariate results for the dyad > stranger contrast and the regions exhibiting greater intersubject correlation (ISC) when subjects listened to dyadic narratives compared to narratives from a stranger. This analysis aimed to identify brain areas that not only showed increased activity for self-relevant content but also displayed enhanced neural synchrony between dyad members when processing such content.

The overlapping regions (Figure 9, Table 4) were identified by masking the univariate results with the significant areas from the ISC analysis. This approach allowed us to isolate the brain regions that consistently demonstrated both heightened activity and increased neural

synchrony for dyadic content compared to stranger-related content. Notably, the overlapping regions were primarily located within the default mode network and attentional/salience networks. Particular areas of the default mode network, such as the OFC, dmPFC, inferior frontal gyrus, and superior temporal gyrus/primary auditory cortex, exhibited both greater activity and enhanced ISC for dyadic narratives. Moreover, regions within the attentional/salience networks, including the middle insula and frontal operculum, also showed overlapping effects.

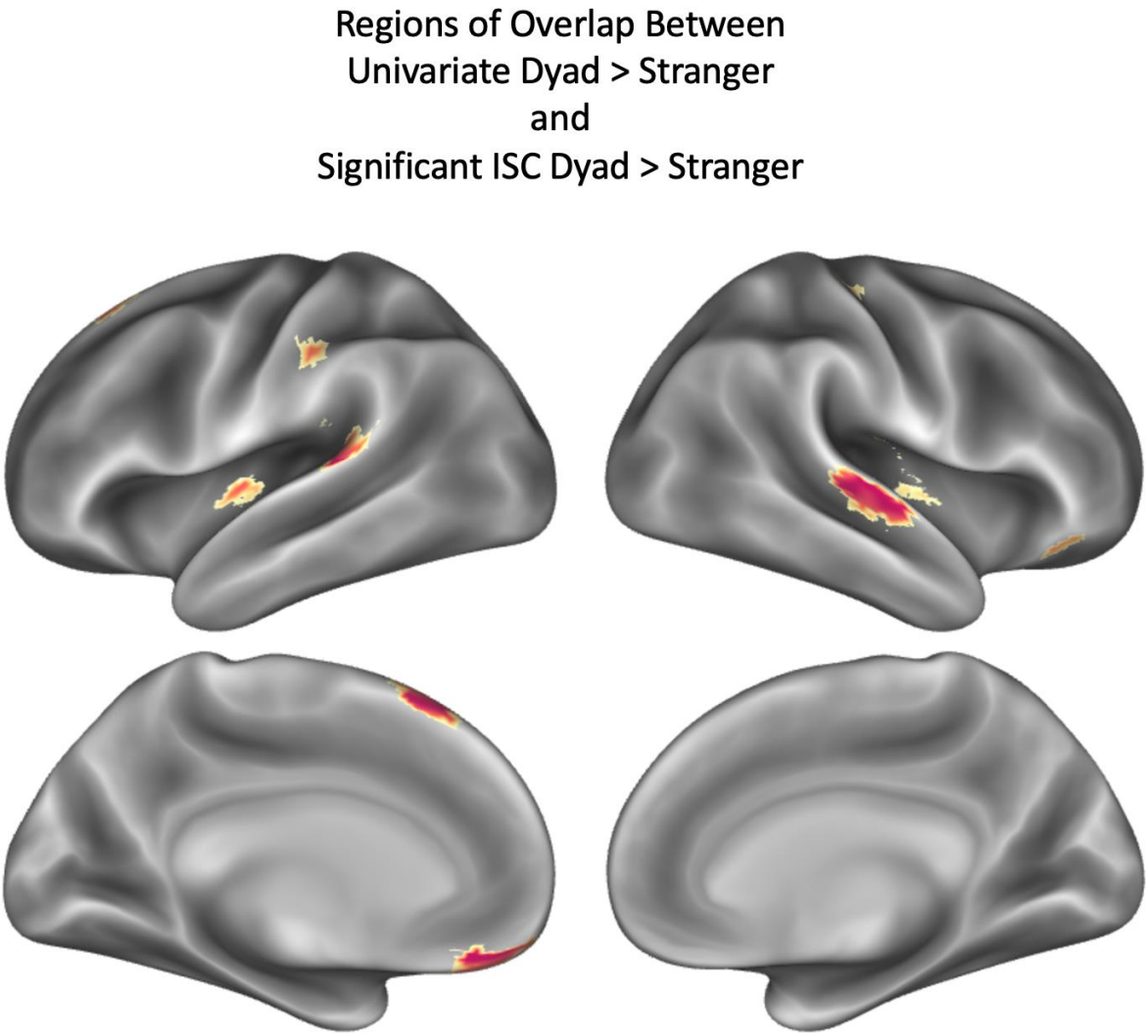

Figure 9. Regions of overlap between univariate dyad > stranger contrast and significant ISC dyad > stranger effects. The brain maps show areas where there was both greater activity for dyadic conditions compared to the stranger condition (based on the univariate analysis) and stronger intersubject correlation for dyadic narratives relative to stranger narratives (based on the mixed effects model). These regions exhibited higher activation and more synchronized brain activity across listeners when processing self-relevant information from oneself or a close other compared to information from an unfamiliar individual.

Table 4

Regions with Overlapping Greater Activity and Greater ISC Synchrony in Dyad vs. Stranger Conditions

MNI Coordinates	Network	Hemisphere	Region	Univariate Z-stat	ISC p-value
60, -11, -2	Default Mode Network	Right	Superior Temporal Gyrus (STG)	9.28	0.039
-3, 43, -21	Default Mode Network	Left	Orbitofrontal Cortex (OFC)	8.91	0.008
-8, 26, 60	Default Mode Network	Left	Dorsomedial Prefrontal Cortex (dmPFC)	8.02	0.009
39, 31, -14	Default Mode Network	Right	Inferior Frontal Gyrus (IFG)	5.40	0.009
-52, -27, 40	Dorsal Attention Network	Left	Superior Parietal Lobule	5.67	0.012
-41, -18, 7	Salience/Ventral Attention Network	Left	Middle Insula	8.65	0.005
43, -13, 3	Salience/Ventral Attention Network	Right	Middle Insula	8.36	0.009
-46, -10, -10	Salience/Ventral Attention Network	Left	Frontal Operculum	8.14	0.004
-65, -49, 24	Salience/Ventral Attention Network	Left	Supramarginal Gyrus (SMG)	3.26	0.018
-64, -33, 12	Somatomotor Network	Left	Posterior Superior Temporal Gyrus	8.46	0.004
35, -26, 53	Somatomotor Network	Right	Postcentral Gyrus	3.62	0.018

Note. These regions represent the overlap between the univariate results from the dyad > stranger contrast and areas with greater ISC for dyad than for stranger. This overlap was obtained by masking the univariate results with the significant regions from ISC.

Discussion

The human brain is highly sensitive to information that holds any type of personal significance and has been shown to engage distinct neural mechanisms when processing this type of information compared to non-relevant information (Meyer & Lieberman, 2018; Northoff et al., 2006; Sui & Humphreys, 2015b; Wagner et al., 2012). Using a novel paradigm that directly compared the neural processing of naturalistic personal narratives that were elicited from oneself, a familiar other, or from an unfamiliar individual, this study extends our understanding of this phenomenon in a more ecological valid framework. Using both univariate and ISC analyses, we provided evidence for the different ways that naturalistic self-relevant information is processed at multiple levels of brain function. Our findings demonstrate that self-relevant narratives engage a wide range of neural networks, including the default mode, language, attention, and salience networks, in a consistent and generalizable way across subjects. This suggests that the processing of self-relevant information involves the coordination of multiple brain systems, with the recognition of personal salience potentially leading to the allocation of attentional resources and enhanced processing at every stage, from low-level sensory processing to higher-order cognitive processing in the default mode network. These results show both increased overall activity from a univariate perspective and also greater neural synchronization across individuals in response to self-relevant narratives compared to those related to unfamiliar others.

At the univariate level, our findings revealed striking differences in brain activation when participants listened to their own narratives compared to those of others. Self-related stories elicited widespread activation across the entire cortical hierarchy, from low-level sensory processing to higher-order cognitive areas. This pattern of heightened activation was present at every stage of information processing, from the initial auditory processing in the thalamus and primary auditory cortex, along the superior temporal sulcus to high order areas in the default mode network (Hasson et al., 2015). This extensive activation observed during self-related narratives, especially at low levels of the sensory processing hierarchy, suggests that these stories are capturing attentional resources, resulting in an enhanced signal-to-noise ratio across multiple levels of processing (Desimone & Duncan, 1995; Hopfinger et al., 2000; Posner & Dehaene, 1994). This interpretation is supported by the recruitment of salience and attention regions that likely serve to signal the importance of self-relevant information and exert top-down influence over the processing of low-level features (Humphreys & Sui, 2016; Menon & Uddin, 2010).

The univariate analysis combining self and close other narratives (dyad) versus stranger narratives revealed notable differences in the neural processing of personally significant information and extended the concept of self-relevance to include not only one's own experiences but also those of individuals with whom we share a meaningful personal connection. When contrasting the dyad condition with the stranger condition, we observed heightened activation in the default mode network for personally relevant narratives. This finding demonstrates that narratives with personal significance engage the default mode network to a greater extent than previously established in studies using non-personal narratives (Hasson et al., 2010; Honey et al., 2012; Simony et al., 2016), highlighting the importance of considering personal significance when investigating the neural processing of naturalistic stimuli.

In contrast, the stranger condition elicited greater activation in limbic system regions such as the amygdala and subgenual ACC and showed virtually no heightened activation in any regions within the default mode network compared to the dyad condition. The heightened engagement of these regions during stranger narratives may be related to an increase in vigilance, uncertainty, and a need for greater emotional regulation when processing unfamiliar or emotionally charged information from an unknown individual (Blackford et al., 2011; Davis & Whalen, 2001; Eippert et al., 2007; Whalen, 2007). This highlights the brain's ability to differentiate between personally significant and non-significant information and provides new insights into the different networks that may be recruited depending on the closeness or familiarity of the other to the individual.

The ISC analysis revealed a similar pattern of results. Although the raw effects plotted across runs suggested that many regions exhibited higher ISC values for stranger narratives compared to dyad narratives, none of these regions reached statistical significance. In contrast, all regions that showed significantly higher ISC or heightened neural synchrony were associated with dyad narratives. Notably, the regions exhibiting higher ISC for dyad narratives included key areas within the default mode network, such as portions of the mPFC, PCC, and TPJ (Alves et al., 2019, p. 201; Buckner & DiNicola, 2019; Mars et al., 2012; Raichle, 2015), indicating that personal significance enhances the inter-subject synchronization of neural responses in brain areas involved in self-referential and social cognitive processes. Furthermore, regions involved in auditory and language processing, as well as salience detection and attentional control (Honey et al., 2012; Menon & Uddin, 2010; Silbert et al., 2014), also demonstrated higher ISC for dyad narratives, suggesting that shared personal relevance facilitates a more consistent processing of the sensory, linguistic, and attentional aspects of the narratives across individuals.

When interpreting the results of the present study, it is crucial to consider the distinct nature of the insights provided by the univariate and ISC analyses, as each approach offers a unique perspective on the brain's response to naturalistic stimuli. The univariate analysis, implemented using a block design, is well-suited for identifying differences in tonic or sustained activation between conditions (Amaro & Barker, 2006; Maus et al., 2010), revealing which networks or regions exhibit heightened activation over prolonged periods throughout the duration of a narrative. This approach provides valuable information about the differential, long-lasting, recruitment of different cognitive processing streams based on the personal relevance of the content. In contrast, ISC analyses capture the consistency of region involvement across participants engaged in the same task (Hasson, 2004; Nastase et al., 2019). By examining the synchronization of activity across participants, rather than overall differences in activation levels, ISC identifies brain regions that exhibit time-locked, stimulus-driven activity across individuals. This approach reveals the specific neural regions within broader networks that are consistently recruited to process the unique characteristics of a given stimulus, providing insight into the generalized neural mechanisms underlying the processing of a particular type of information. Together these approaches compliment each other by showing where in the brain there is differential systemic allocation of neural resources throughout the duration of the narrative and also where, within these large networks, there are focal regions in which a particular type of information is being processed in a consistent and generalizable way across subjects.

The convergence of univariate and ISC findings in specific brain regions sheds light on the neural mechanisms that underlie the processing of personally significant information. The bilateral middle insula emerged as a key region, exhibiting both heightened activation and increased synchronization across participants during dyad narratives. This finding suggests that

the insula plays a crucial role in integrating personally relevant information, possibly by mediating the interaction between self-referential processing and the allocation of attentional resources (Conway et al., 2016; Kurth et al., 2010; Menon & Uddin, 2010; Perini et al., 2018; Uddin, 2015). This attentional enhancement is evident in the superior temporal gyrus, a region implicated in auditory and language processing, which demonstrated the highest increase in both activation and ISC for dyad narratives. The heightened activity in this sensory processing region, coupled with the increased synchronization across individuals, indicates that personal significance not only amplifies the neural response to the incoming stimuli but does so in a consistent and generalizable way across participants (Silbert et al., 2014). Furthermore, the engagement of the OFC and the dmPFC, both regions within the default mode network, underscores that these findings are apparent at all levels of the cortical hierarchy leading to not only enhanced sensory processing but also heightened high level social cognitive processing as well (Buckner & DiNicola, 2019; Hasson et al., 2015).

One limitation of the present study is the variability in the nature and strength of the relationships between dyad members. While all participants were familiar with their dyadic partners, the specific type of relationship varied, including friends, romantic partners, and work acquaintances. This heterogeneity in relationship types could potentially influence the degree of neural synchrony observed, as different types of relationships may be associated with varying levels of shared experiences, knowledge, and emotional connection (Brown et al., 2022; Parkinson et al., 2018; Puusepp, 2023). Consequently, the results might differ if the relationships within the sample were more homogeneous. However, it is important to note that this variability in relationship types was an intentional design choice which allows for future investigations into the potential effects of relationship type on neural synchrony using this dataset and maximizes

generalizability to various kinds of relationships. Despite the increased noise introduced by this heterogeneity, the fact that our main hypotheses regarding heightened default mode network involvement for personally relevant narratives were supported suggests that the observed effects are robust and generalizable across different types of familiar relationships. Nevertheless, future research could benefit from exploring the influence of relationship type on neural synchrony more directly by comparing dyads with specific types of relationships or by recruiting a more homogeneous sample of dyads.

Another potential limitation of the current study is the variability in the content and emotional valence of the personal narratives used as stimuli. Participants were asked to share stories covering a range of experiences, including positive and negative past events, positive and negative future possibilities, and personal ideologies. These different types of narratives may elicit varying degrees of personal relevance and emotional responses, which could potentially impact the results (Holland & Kensinger, 2010; Svoboda et al., 2006). For example, a highly emotionally charged negative past event might lead to stronger neural synchrony compared to a more neutral personal ideology, regardless of the familiarity of the speaker. It is important to note though that the narrative prompts were deliberately chosen to capture a wide range of personally meaningful information about each individual, drawing upon the life narrative interview approach developed by McAdams (2001). This approach aims to elicit a comprehensive understanding of an individual's life story, including their significant experiences, future aspirations, and guiding beliefs. The use of a diverse set of narrative prompts was done intentionally to obtain a more holistic representation of each person's identity and personal relevance. Despite the potential confounds introduced by the variability in narrative content and emotional valence, the fact that our main findings were consistent across the

different types of narratives suggests that the effects of personal relevance on neural processing are robust and generalizable. However, future studies could explore the influence of narrative content and emotional valence more systematically by controlling for these factors or by directly comparing neural synchrony across different types of narratives.

A final notable limitation of the present study is the potential influence of frame of reference on neural synchrony. In our experimental design, all self-narratives were presented in the first person, while familiar other and stranger narratives were presented in the third person. Additionally, participants were inherently more familiar with the details of their own stories compared to those of their dyadic partner or the stranger. This discrepancy in frame of reference and familiarity could potentially affect intersubject correlation, particularly if one member of the dyad recalls more details than the other while listening to the same narrative (Smirnov et al., 2014; Zadbood et al., 2017). If frame of reference were a primary driver of neural synchrony though, we would expect to observe higher ISC for the stranger narratives, where both members of the dyad were listening to a third-person account, compared to the self/partner ISC, where one person was listening in the first person while the other was listening in the third person. However, our findings revealed the opposite pattern, with significantly higher ISC for self/partner narratives compared to stranger narratives. This suggests that the effects of personal relevance on neural synchrony are strong enough to overcome any potential influence of frame of reference, highlighting the importance of considering the social and emotional significance of stimuli in the study of brain responses to naturalistic stimuli.

Summary

These findings provide a more comprehensive understanding of the neural dynamics that support the processing of personally significant information and emphasize the interplay between self-referential, attentional, and linguistic processes in shaping the brain's response to narratives that hold personal relevance. The present study highlights the importance of considering personal significance in the design and interpretation of experiments employing naturalistic stimuli, as the neural processing of such stimuli is heavily influenced by the personal relevance of the content. By demonstrating the differential engagement of large-scale brain networks in response to self-relevant narratives, our results offer novel insights into the mechanisms underlying the integration of personal significance with real-world, ecologically valid information. Future research should further investigate the role of personal relevance in shaping neural dynamics during naturalistic stimulation, as well as its potential implications for cognitive processes such as attention, memory, cognitive control, and social cognition.

Chapter III

Normativity vs Uniqueness:

Effects of Social Relationship Strength on Neural Representations of Others

This work was co-authored with Robert S. Chavez. I was the lead author of the publication and responsible for the writing of the manuscript with edits provided by the co-authors. I led the methods and analyses with input and assistance from the co-authors.

Introduction

Social cognition is anchored heavily on our ability to infer the intentions, character, and preferences of another's mind, despite being fundamentally blind to the true source of this information. Initially, in the absence of direct experience or prior interaction, our brains default to relying on broad, categorical assumptions to navigate potential interactions with new acquaintances (Biesanz, 2021a; Chan et al., 2011a; Rogers, 2021). Such early interactions are often overshadowed by stereotypes and generalizations, which effectively strip the individual of their unique identity. However, it is through deliberate and effortful engagement that a transformative process begins to unfold. This process gradually dismantles biases, allowing individuation and nuanced understanding to enrich the mental model we construct of that person (Haslam, 2006; Swencionis & Fiske, 2013). How though, does this transition from generalized perceptions to nuanced individualized understanding manifest in the brain?

In this study, we test the hypothesis that the strength of social relationships directly influences how the brain represents others, leading to increasingly individualized perceptions within regions of the brain critical for person perception and social cognition more broadly. This hypothesis builds on the premise that familiarity and emotional closeness not only deepen our understanding of others (S. T. Fiske & Neuberg, 1990) but may also modify the neural underpinnings of this perceptual process. Our research diverges from directly testing the out-

group homogeneity effect (Tajfel, 1970), aiming instead to test a novel but related hypothesis that gradients of individualization may exist within pre-established social networks. Specifically, we examine whether the strength of social ties within a group of acquaintances can predict the extent to which a neural representation of a specific group member deviates from an averaged 'normative other' activity pattern.

To test the theory that social relationships influence the creation of more unique brain representations, we employed the use of a round-robin design. Investigating true relationships in cognitive neuroscience is crucial, as it provides deeper insights into the cognitive processes that are unfolding in real-world social interactions. Traditionally, research in person perception has often relied on the use of stimuli that describe fictional or socially distant individuals (Hassabis et al., 2014; Kelley et al., 2002; Tavares et al., 2015), leaving room for potential limitations in the ecological validity of the findings. Additionally, studies that have used subjects' own close others as stimuli (Chavez et al., 2017; Thornton & Mitchell, 2017) have encountered challenges in controlling for unique variables inherent in each relationship, such as the degree of closeness, similarity, and the history of interactions, which can vary widely across participants. The round-robin design offers a robust solution to these challenges by involving small groups of individuals who are already familiar with each other to varying extents (Chavez & Wagner, 2020; Guthrie et al., 2022). This method allows researchers to not only probe the cognitive responses of perceivers in a more controlled and authentic context but to also demonstrate how these perceptions are influenced by the varying dynamics within the group relationships. As a result, this approach allows for an investigation of genuine interpersonal relationships, offering valuable insights into the cognitive processes underpinning person perception.

Influential behavioral theories on impression formation highlight our tendency to rely on generalized or normative information when understanding others (Biesanz, 2021a; Cronbach, 1955; S. T. Fiske & Neuberg, 1990; K. H. Rogers, 2021). This approach is advantageous when meeting someone new, given that, statistically, a random person is much more likely to resemble the average person than any specific known other (Chan et al., 2011a; K. H. Rogers, 2021). While much of the research has focused on the mechanisms of impression formation, certain studies have explored how our dependence on normative profiles is affected by the depth of social relationships or with the passing of time (Biesanz et al., 2007; Human et al., 2020). These studies, however, have led to conflicting results with some showing a decrease in the use of normative profiles over time or as relationships emerge and others showing an increase. Furthermore, it has been suggested that specific social goals encourage us to differentiate individuals (S. T. Fiske & Neuberg, 1990), for reasons ranging from enhancing group identity to fostering trust (S. T. Fiske et al., 1999). Despite this rich landscape of behavioral insights, a gap remains in our understanding of how the brain balances normative and individuated representations in the context of others, particularly in how this may be influenced by the strength of social relationships.

Much of the research that has extended this social psychological work on person perception into the neural domain has focused on impression formation and the role of stereotypes in categorizing others (Barnett et al., 2021; Brooks & Freeman, 2019; Freeman et al., 2012; Quadflieg et al., 2011; Stolier & Freeman, 2016). This work, along with research on person perception and social cognition more broadly, has led to the discovery of a broad network of regions that collectively allow the brain to process perceptual features of a particular other while also allowing for the inclusion of social information in the perceptual process (Brooks &

Freeman, 2019; R. Saxe & Kanwisher, 2003; Wagner et al., 2012). Notably, it has been demonstrated through the use of multivoxel pattern analysis that generalized information, such as gender or race, can be detected within this network (Brooks & Freeman, 2019; Contreras et al., 2013; Stolier & Freeman, 2016). However, additional studies suggest that stimulus identity can also be decoded within these socially oriented brain regions, indicating that the brain can also utilize a unique representational structure when forming perceptions of others (Axelrod & Yovel, 2015). This dual aspect of brain function within this brain network underscores the social cognitive capacity to move from generalized to unique perceptions of others depending on the context of the situation.

Round-robin designs are uniquely positioned to advance our understanding of normative profiles in the brain. It enables the modeling of multiple representations of known others within each subject and permits the analysis of how various types of relationships might influence the differences in these representations (Guthrie et al., 2022). By utilizing the extensive data on known others represented in each subject's brain throughout the study, this method enables the generation of normative brain patterns through the aggregation of these individual representations. This approach may provide a more accurate reflection of how generalized representations are used in real-world settings. Additionally, unlike studies that focus on perceiving a single known individual (Chavez et al., 2017), this design can differentiate whether subjects are using generalized or unique cognitive representations based on their relationship with each target, offering a more nuanced understanding of how these brain regions are processing social information.

To this end, the current study utilized fMRI to investigate the hypothesis that greater levels of social relationship strength would result in more distinct neural multivoxel patterns in

socially oriented brain regions. By employing a round-robin interpersonal perception paradigm, we were able to collect behavioral and neuroimaging data from 20 pre-existing groups of 5-6 people from various real-world social networks. The study design allowed us to create an averaged 'normative other' brain representation from brain responses each subject exhibited while thinking of each of their group members across the entire data set. This normative pattern was then used to test whether social relationship metrics could predict the degree to which a subject was using either an average representation of a particular other or whether they had formed a more unique multivoxel pattern of activity to represent their group member. This approach, combined with the behavioral measures of social relationship strength, allowed us to examine whether gradients of individualization exist within pre-established groups and whether social relationships modulate the way that our brains model these perceptions.

Methods

Participants

Using a multi-group round-robin design, we recruited a total of 120 right-handed participants between the ages of 18 and 51 (48 females and 72 males) from 20 pre-existing independent social network groups, including student organizations, local businesses, and friend groups. Participants ranged in age from 18 to 51 with a mean age of 23.5 and standard deviation of 7.2. A majority of participants identified as Caucasian (77.9%), 4.4% identified as Black or African American, 2.7% identified as Asian, 1.8 % identified as American Indian or Alaska Native, 1.8% identified as Native Hawaiian or Pacific Islander, 9.7% identified as more than one race, and 1.8% chose not to identify their race. A majority of participants identified as non-Hispanic (84.1%). Six people were recruited from each of the 20 groups, and all participants

within each group were familiar with one another but had various degrees of closeness with the other members.

Procedures

Behavioral

Across all groups, five participants failed to fulfill scheduled experimental session appointments, two subjects were excluded due to unusable imaging data, and two subjects were excluded for missing or incomplete behavioral data. This left a final N = 111 subjects in total with at least five participants per group. All participants were screened for MRI contraindications and had normal or corrected-to-normal vision. Each participant took part in two sessions. The initial session focused on behavioral ratings, during which participants completed a set of questionnaires for themselves and their fellow group members. The second session was an MRI scanning session, where participants made trait judgments for both themselves and their peers while undergoing functional neuroimaging. Participants provided informed consent in accordance with the guidelines established by the Internal Review Board at the University of Oregon for each session and received compensation for their involvement after completing each part of the study.

Participants were brought into the lab for the first session and tasked with responding to a series of inquiries about themselves and a specific group of known peers. Each participant was requested to provide ratings of closeness and similarity towards each individual in the group. All responses were logged using PsychoPy stimulus presentation software (Peirce et al., 2019). During the behavioral assessment, participants viewed a range of scales (ratings 1–5), with either

one of their peers' names or their own displayed above each scale. A single question was presented at the top of the screen, prompting participants to rate their peers or self-assess on that particular characteristic. In order to ensure comprehensive feedback, participants had to rate every peer before proceeding to subsequent questions. The entire session took approximately 1 hour.

The ratings gathered during this session were then utilized to compute the behavioral interpersonal relationship evaluations for subsequent fMRI studies. These assessments were derived from self-reported responses on a 1–5 Likert scale related to various statements about friendship, knowledge of the individual, affinity, and perceived similarity. These responses were combined into an overall measure of social relationship strength. This approach mirrored the one employed by Guthrie et al. (2022) and was reassessed for its application in the normative analyses.

Neuroimaging

In a separate lab session, the participants returned to complete the fMRI part of the study. While in the scanner, they were tasked with a common trait-judgment activity that has been widely used in research on self and other processing (Kelley et al., 2002; Mitchell et al., 2011). The study employed a comprehensive round-robin design, meaning each participant acted as both an evaluator and target stimulus for every other participant. A screen displayed two words arranged vertically in white text on a black background. For each trial, the top word showed either “Self” or one of five group members' names from their own group. During an initial session outside of the scanner, participants indicated which name they most frequently used when referring to each group member. These names were then utilized during fMRI sessions

(e.g., “Alexander” might have been changed to “Alex” or "Zander”). This was done to ensure that participants did not need extra time inside the scanner to try to figure out who each target was.

The bottom word presented one of 60 valence-balanced adjectives (e.g., “Happy”, “Clumsy”, and “Smart”; (Anderson, 1968)) for 2000 ms followed by fixations lasting 2000–12,000 ms interspersed with intermittent passive fixation trials. Jittered trials were optimized using Opseq2 (Dale, 1999). Participants had to indicate whether a given trait adjective described themselves or one of their group members by pressing buttons labeled “yes” or “no”. All targets appeared in each run with a total of twelve trials per target per run where all targets were paired with these same twelve trait adjectives. The same targets were then used in each subsequent run but were then paired with a new set of 12 trait-adjectives in each one, resulting in all targets being paired with all 60 trait adjectives over the course of the experiment. Individual traits were only presented once per target randomly across all runs in the experiment. No two participants were presented with the same target/trait-adjective order across the experiment to account for potential order effects.

MRI was conducted with a Siemens 3T Skyra scanner using a 32-channel phased array coil. Structural images were acquired using a T1-weighted MP-RAGE protocol (175 sagittal slices; time repetition [TR]: 2500 ms; time echo [TE]: 3.43 ms; flip angle: 7°; 1-mm isotropic voxels). Functional images were acquired using a T2*-weighted echo planar sequence (TR: 2000 ms; 72 axial slices; TE: 25 ms; flip angle: 90°; 2-mm isotropic voxels). For each participant, we collected five runs of the round-robin task (188 whole-brain volumes per run). In order to correct for distortion due to B0 inhomogeneity, we also acquired a field map (TR: 6390 ms; TE: 47.8

ms; effective echo spacing: 0.345 ms). The total length of time for the entire scanning session was approximately 75 min, and each of the five functional runs was approximately 6 min long.

Analysis

Preprocessing

Functional imaging data acquired for the fMRI task underwent preprocessing and the estimation of voxel responses was completed using FSL (S. M. Smith et al., 2004). The data went through mean-based intensity normalization, high-pass filtering (Gaussian-weighted least-squares straight line fitting with sigma = 100 s), and spatial smoothing with a 4-mm FWHM Gaussian smoothing kernel. A multistep normalization procedure was employed to register the results to standard space. First, functional data were corrected for spatial distortion by using a field map unwarping before aligning functional data to each participant's anatomical scan by using boundary-based registration (Greve & Fischl, 2009) in conjunction with a linear registration with FSL's FLIRT tool. These images were then warped to a 2-mm Montreal Neurological Institute template by using nonlinear registration with FSL's FNIRT tool and a 10-mm warp field. All first-level task-specific analyses were initially conducted in native space before being warped into standard space for final analyses. Parameter estimates were independently calculated for each of the five group members within all five runs completed by each participant. These responses were then combined in a second level within-subject fixed-effects analysis yielding parameter estimates for each of the five group members' "target" conditions. Although a self-condition was collected during this experiment, none of the self-estimates were used in this particular analysis. All estimates used in the analysis represented

perceivers thinking of other group members. Normalized (i.e., z-score) voxel responses for each condition were extracted from a set of 400 parcels from the Schaefer atlas (Schaefer et al., 2018).

Normative Other Estimation

The result of the response pattern estimation process was a set of 5 estimates (one for each perceived target) in each of the 400 parcellated regions of interest (ROI) for every one of the 111 subjects. Across all subjects included in the analysis, there were 555 unique estimates of cognition linked to a perceiver thinking about a known other. Parameter estimates within each ROI were flattened into 1D response vectors to allow for alignment of responses across subjects. Similar to the procedure used when calculating a normative personality profile (K. H. Rogers, 2021), the 1D response vectors for all unique estimates in each ROI across all of the subjects were averaged to allow for the calculation of a normative other estimate. This estimate then became the anchor of comparison in each ROI for calculating the dissimilarity between a specific representation of a particular other and the normative averaged response of all others in the study. It is important to note that the particular target that was being compared to the normative average was systematically left out of the calculation of the normative average in each iteration of the analysis to avoid a potential biasing of the correlation distance between that target's particular representation and the norm. This meant that none of the estimates of the perceiver thinking about that target were included in the average as well as any of the estimates created from other group members thinking of that same target. A schematic of this approach is shown in Figure 10.

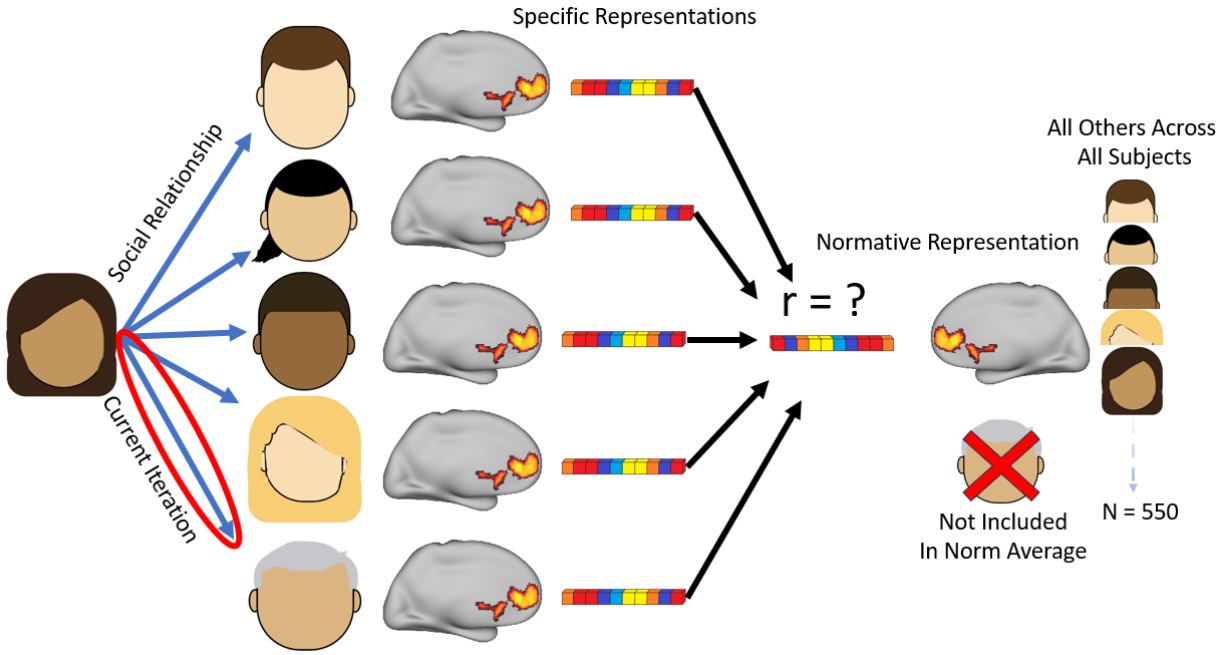

Figure 10. Illustration of the fMRI analysis process for assessing unique and normative representations of group members. Each subject was scanned while performing a trait adjective task designed to engage cognitive representations of each of their group members. Brain activity estimates were acquired for each group member and then segmented into 400 brain regions. For each region, the brain activity from all subjects thinking about each of their group members, except for the specific representation being compared, was averaged to create a normative other activity pattern (N = 550). The Spearman rank correlation distance between a subject's specific representation of a particular group member and the normative average was calculated for every region. This distance served as a measure of dissimilarity, with the strength of the social relationship between the subject and each group member used to predict this dissimilarity across regions.

Distance Between Specific Others and Normative Average

Consistent with previous studies examining the similarity/dissimilarity of neural representations (Kriegeskorte, 2008), we computed, within each subject, the dissimilarity between neural response patterns of specific peer response vectors from the averaged normative other response vector by using Spearman rank correlation distance. Within each parcel, we calculated correlation distances between a subject's representation of a particular other and the

normative other estimate and this was done iteratively for all five group members that that subject perceived in the experiment. These correlation distances were then related to social relationship scores that were obtained in the behavioral session in which each perceiver rated the other members in their group. We sought to determine whether the distance in multivoxel similarity between each particular other and the normative average was related to social relationship strength. In order to account for the nested group structure of the comparisons, a linear mixed effects model was employed in which the perceiver was modeled as a random intercept nested within each group. False discovery rate correction was applied to correct for multiple comparisons across the 400 parcel ROIs.

Trait Endorsement Analysis

In addition to the neuroimaging data, we also analyzed the trait endorsement ratings collected during the scanner session to investigate whether the strength of social relationships influenced the similarity between an individual's trait endorsements for a specific target and the normative endorsement pattern derived from all other subjects rating other targets. By comparing the results of this endorsement analysis with the findings from the neuroimaging data, we aimed to gain a more comprehensive understanding of how social relationships shape both explicit behavioral responses and implicit neural representations in the context of person perception. For each subject, we extracted the endorsement ratings indicating whether they responded “yes” (the trait described the target) or “no” (the trait did not describe the target) to the trait adjectives paired with each target in each run.

To perform an analysis like the one conducted on the brain data, we first sorted the dataset in alphabetical order by trait word to ensure endorsement ratings were matched across

targets and subjects. For each target, we isolated their endorsement ratings and calculated a normative endorsement by averaging the endorsement ratings for all other subjects rating all other targets, excluding the specific target being analyzed. We then computed the Euclidean distance between the target's endorsement vector and the normative endorsement vector. This process was repeated for all possible combinations of targets and subjects across the dataset.

To investigate the relationship between social relationship strength and the similarity of endorsement ratings to the normative pattern, we extracted the social relationship scores for each subject rating the particular target used in the analysis. These scores were then used as predictors in a linear mixed effects model, with the Euclidean distance values as the dependent variable. The model included subject nested within group as a random effect to account for the hierarchical structure of the data.

Code and Data Availability

Data, materials, and code used to generate the findings in this study are openly available via an Open Science Framework repository at: https://osf.io/wnyma.

Results

Neuroimaging Results

Participants were recruited from real-world social groups in which each person had previously established relationships with all other members within each round-robin group. Despite the previous acquaintanceship, there remained a high degree of social relationship strength variability across all participants (M = 3.46, SD = 0.83). Moreover, variability in social relationship strength was observed both within groups and between groups, with some groups

having higher average social relationship strength than others (Figure 11). Means and distributions of social relationship strength for each group are shown in the density ridgeline plots in Figure 10. Thus, as expected, these groups of previously acquainted individuals tended to have higher social relationship strength than would be expected of strangers, but the variability needed to model the fMRI similarity metrics was present in all groups.

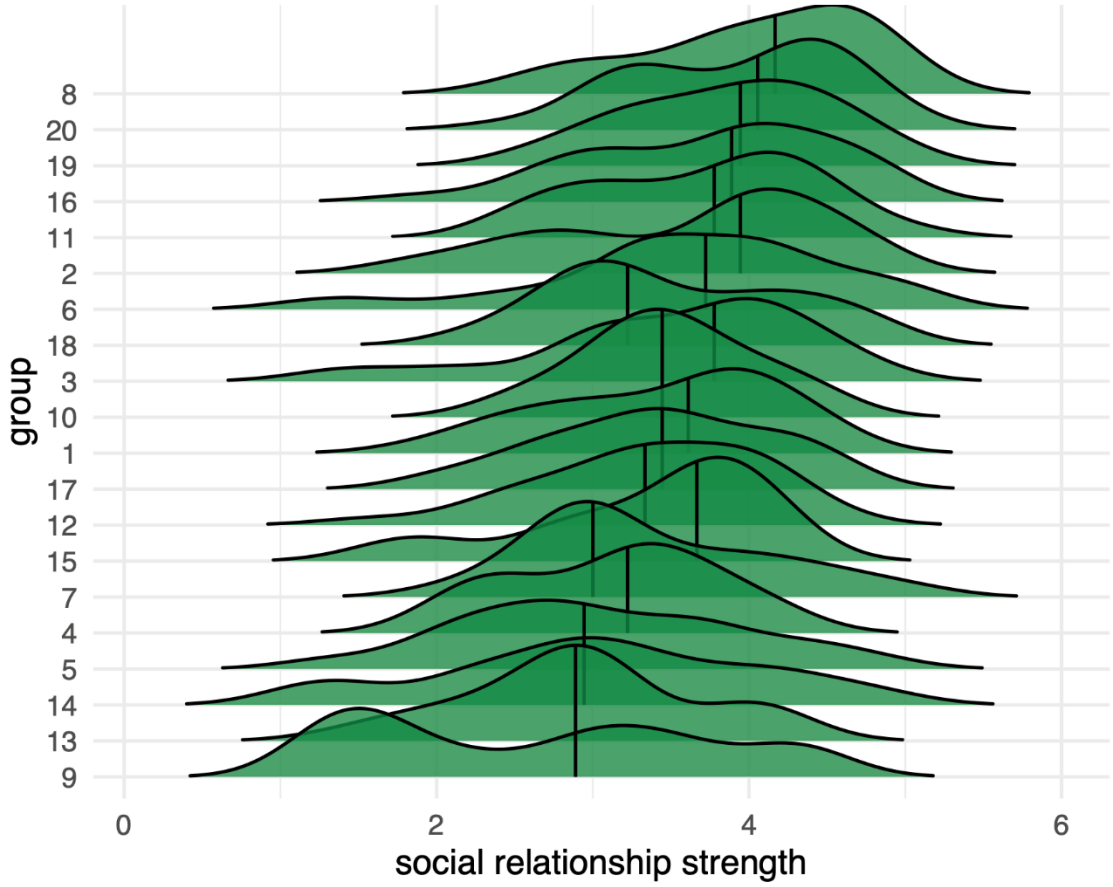

Figure 11. Variability in social relationship strength across 20 different groups. The ridgeline plot illustrates the distribution of perceived relationship strength within each group, highlighting both within-group variability and between-group differences. Groups exhibit a range of relationship strengths, with some groups reporting stronger ties and others weaker connections. Groups were deliberately recruited to enhance variability, including friend groups, academic peers, and work acquaintances, to capture a wide spectrum of familiarity.

All fMRI analyses were performed in the same manner within each of the 400 parcels independently, and significance values were corrected for multiple comparisons using FDR. We hypothesized that greater social relationship strength would be reflected in more unique (i.e., lower degree of similarity) multivoxel pattern similarity between an individual target and the normative other. The results of these analyses are displayed in Figure 12. Consistent with this hypothesis, social relationship strength between a perceiver and a target was significantly associated with a greater degree of representational uniqueness within brain regions implicated in social cognition. These regions include the left vmPFC ($b = 0.006$, SE = 0.004, $p < .001$), the right vmPFC ($b = 0.005$, SE = 0.001, $p = 0.003$) and the left ventral anterior insula ($b = 0.006$, SE = 0.001, $p = 0.010$). A complete list of all significant regions from these analyses is summarized in Table 5.

Table 5

Brain Regions with Unique Multivoxel Patterns Predicted by Social Relationship Strength

Brain Region	Peak Coordinates (x, y, z; MNI)	Estimate	Standard Error	P-value (FDR Corrected)
Left vmPFC	(-6,45,9)	0.006	0.004	<.001
Left Ventral Anterior Insula	(-34,16,-9)	0.006	0.001	0.010
Left dlPFC	(-27,55,12)	0.003	0.0009	0.050
Right vmPFC	(7,42,5)	0.005	0.001	0.003
Right Middle Temporal Sulcus	(62,-41,-9)	0.003	0.0009	0.021

Note. This table lists brain regions where the strength of social relationships was found to significantly predict the uniqueness of multivoxel pattern activity, as indicated by greater dissimilarity from the normative average. Peak coordinates are presented in MNI space. P-values are FDR-corrected for

Table 6

Brain Regions with Generalized Multivoxel Patterns Predicted by Social Relationship Strength

Brain Region	Peak Coordinates (x, y, z; MNI)	Estimate	Standard Error	P-value (FDR Corrected)
Left PCC/Precuneus	(-5,-61,30)	-0.003	0.0009	0.012
Left Precuneus	(-5,-65,52)	-0.003	0.0009	0.039
Left dlPFC	(-20,61,8)	-0.003	0.0009	0.050
Left Intraparietal Sulcus	(-44,-42,47)	-0.002	0.0007	0.039

Note. This table lists brain regions where the strength of social relationships significantly predicted the generalization of multivoxel pattern activity, as indicated by greater similarity to the normative average. Peak coordinates are presented in MNI space. P-values are FDR-corrected for multiple comparisons.

Although our hypothesis predicted that stronger social ties would be associated with greater uniqueness in representations of specific others, some regions showed an inverse relationship. In these areas, higher social relationship strength was associated with representations of specific others that were more closely aligned with the normative average (higher degree of similarity). The observations in the left PCC/precuneus region fall within an area known to be involved in social cognition. This included a more ventral region that overlaps with the PCC and precuneus (b = -0.003 , SE = 0.0009, p = 0.012) and a more dorsal region within the precuneus (b = -0.003 , SE = 0.0009, p = 0.039).These results are also shown in Figure 12, and a complete list of all significant cluster statistics for the individual target results can be found in Table 6.

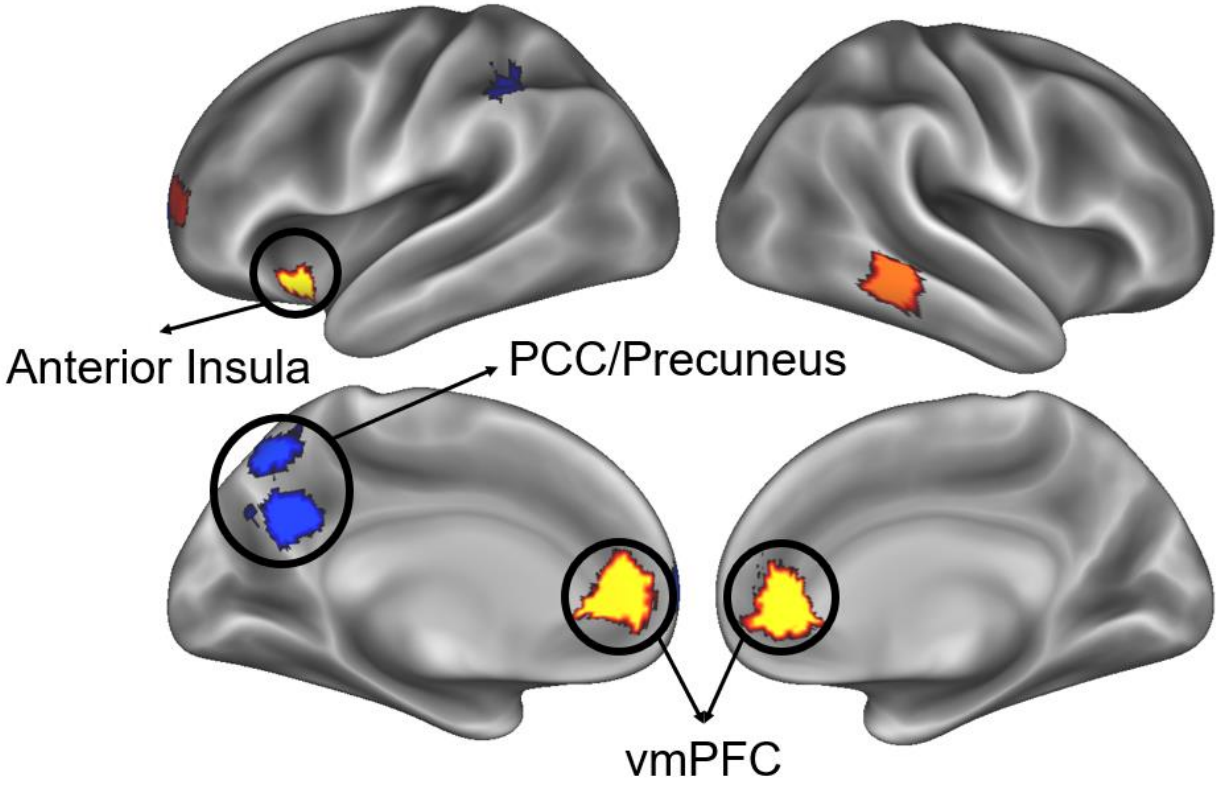

Figure 12. Brain regions demonstrating significant relationships with social relationship strength in terms of neural representation uniqueness and generalization for a specific group member. Areas highlighted in warm colors (red-yellow) indicate regions where stronger social ties predict greater dissimilarity between an individual's neural representation of a group member and the collective normative average, signifying a more unique representational structure. Conversely, areas in cool colors (blue shades) mark regions where stronger social relationships correspond to more generalized neural patterns, aligning more closely with the normative group average, indicating an inverse relationship to the predicted effect.

Trait Endorsement Results

To investigate the relationship between social relationship strength and the similarity of trait endorsements to the normative pattern, we conducted a linear mixed effects analysis. The model included social relationship scores as predictors of the Euclidean distance between a target's endorsement vector and the normative endorsement vector, with run nested within subject, further nested within dyad, as random effects.

The analysis revealed a significant effect of social relationship strength on the similarity of trait endorsements to the normative pattern ($\beta = -0.03$, $SE = 0.004$, $t = -6.83$, $p < .001$). Contrary to

our expectations, the negative coefficient indicates that as social relationship strength increased, the Euclidean distance between a target's endorsement vector and the normative endorsement vector decreased. In other words, stronger social relationships were associated with more normative or average trait endorsements.

Discussion

Navigating the diverse connections of our daily social experience necessitates the flexible adaptation of perceptual strategies to effectively model our representations of others in accordance with the nature of our relationship with them. Generalized cognitive representations may serve as efficient heuristics for forming impressions of acquaintances or strangers (Chan et al., 2011a; K. H. Rogers, 2021), but with close others, it becomes increasingly important to capture their nuanced individuality in the mental models we form of them (Swencionis & Fiske, 2013). Using a round-robin fMRI design, the results from this study demonstrated that social bonds have a modulatory effect on the ways in which neural representations of close others are formed. By establishing a 'normative other' activity pattern, aggregated from the collective representations of others across the study, and contrasting individual representations against this norm, we show that the strength of social relationships predicts the uniqueness of neural representations of others in key regions involved in social cognition.

The vmPFC plays a pivotal role in the process of person perception, as it has been shown to be preferentially engaged in the high-level abstract judgements required for making attributions about enduring traits (Harris et al., 2005; Ma, Baetens, Vandekerckhove, Kestemont, et al., 2014; Mitchell, 2004; Schiller et al., 2009; Van Overwalle, 2009). Neuroimaging and virtual lesion studies have consistently shown that the vmPFC appears to be necessary and also distinctly recruited when making these types of trait judgements irrespective of general valence

or the interpretation of specific behaviors in social stimuli (Kestemont et al., 2016; Ma, Baetens, Vandekerckhove, Kestemont, et al., 2014; Ma, Baetens, Vandekerckhove, Van der Cruyssen, et al., 2014a; Van Overwalle, 2009). This specificity suggests that trait information is not processed in a distributed manner across the networks involved in social cognition but rather that the vmPFC is a central component in the brain's ability to create these types of representations. Furthermore, it has been demonstrated that the univariate activity patterns in the vmPFC increase linearly with the similarity and closeness of a social target being judged, highlighting that this trait space is likely utilized more with close or familiar others (Krienen et al., 2010; Mitchell, Macrae, et al., 2006).

Our study further elucidates the role of the vmPFC, revealing that it not only develops distinct neural activity patterns to represent others, but that these representations become increasingly more unique as the social relationship with the target increases. The distinct role of the vmPFC in creating personalized trait representations is crucial for the interpretation of these findings as the analyses were performed in dissimilarity space. Understanding that the vmPFC is likely engaged in computing high-level trait information allows us to infer that the normative pattern that we've created represents an average trait space rather than an aggregate of neural responses to diverse cognitive processes such as attention or sensory processing. This suggests that the activity patterns observed in the vmPFC for specific targets diverge from this normative trait space and likely represent the nuanced ways in which we represent the specific traits of close others.

Another crucial component of creating and maintaining close relationships is the involvement of emotion. It has been demonstrated that along with the high-level trait attributions that the vmPFC is involved in, it has also been shown to represent affective meaning (Roy et al.,

2012; Rudebeck et al., 2008). This is separate from general valence in that it requires an integration of valence signals with higher level conceptual and contextual elements. Patients with lesions to the vmPFC have been shown to routinely exhibit irregular physiological and appraisal behavior in response to personally relevant social and emotional information (Beer et al., 2003; Damasio et al., 1990; Hornak et al., 2003; Leopold et al., 2012). Parallel to the findings in the vmPFC, we also found evidence for the unique representational structure effect in the anterior insula. This region is also heavily involved in the integration of emotion with higher-level subjective and social information (Chang et al., 2013; Craig, 2011; Kurth et al., 2010; Yoon et al., 2021). Previous computational work also demonstrated the role of the anterior insula in social decision making related to the creation of social alliances (Lau et al., 2020). Together, these findings demonstrate a possible individuation of trait information and affective meaning in these anterior social regions as social ties become more prominent with the individual that is being assessed.

It is widely recognized that the brain implements different learning strategies based on the context and demand of the situation (Schapiro et al., 2017). It is highly efficient for the brain to implement the use of prototypical heuristics to navigate scenarios that involve generalizable features and the creation of these mental models likely involves the extraction of normative and consistent information that is routinely encountered (Bowman et al., 2020). To successfully navigate novel situations however, it is also important to differentiate salient information that may be useful above and beyond the predictive power that a generalized prototypical model can afford. It is important to note, however, that the brain has been shown to be capable of creating and maintaining both types of representational structures simultaneously and in parallel (Schapiro et al., 2017). Discriminative attention has been proposed as a key factor in shifting the

reliance away from a generalized representation by instead initiating a process that accommodates new information into the existing models (Kahneman et al., 1983; Musslick et al., 2020). Forming and maintaining strong, meaningful social bonds likely involves applying these discriminative attentional processes to recognize and appreciate the unique qualities that distinguish individuals from the broader collective (S. T. Fiske et al., 1999).

While we observed the anticipated unique representation effect in certain anterior areas within this network, such as the vmPFC and anterior insula, our findings also revealed an inverse effect in other regions also known to be involved in social cognition. Notably, posterior regions, including parts of the PCC and precuneus, displayed patterns where stronger social ties resulted in a representational structure that was more generalized, converging more with the normative average. This divergence in neural processing between anterior and posterior regions suggests that the social networks in the brain do not operate uniformly. Instead, these anterior and posterior regions which are often shown to be coactive in univariate studies (Andrews-Hanna et al., 2014; Mak et al., 2017; Raichle et al., 2001; Yeshurun et al., 2021), are likely engaging in separate and distinct computations when representing others. This highlights the role of the complimentary learning systems approach and may be evidence for both generalized and individualized representations being maintained and utilized in parallel by different components of the network.

Although there was evidence found for other regions showing the same pattern of unique brain activity being predicted by social relationship strength and also other regions showing the inverse, these regions were outside of the hypothesized goal of the study and should be interpreted with caution. The analysis does indeed show that the patterns of activity in these regions are deviating from or converging with an average when thinking about close others, but

it is unclear what that average or what the resulting idiosyncratic or generalized activity is computing. The brain is highly entangled, and it is likely that a shift in representational structure in the anterior social regions that we observed has a driving effect on other regions that may be close by or functionally coupled with these regions in ways that we have yet to understand (Chavez, 2021). Additionally, these findings may suggest broader network changes in the way the brain orients itself to self-relevant information in general as close others are more likely to activate regions involved in the allocation of discriminative attention (Humphreys & Sui, 2016) and the instantiation of long term and working memory (Northoff et al., 2006; Sui & Humphreys, 2015a; Wagner et al., 2012). Although these unexpected findings offer intriguing insights into the complexity of neural processes, further research is needed to fully understand the significance and mechanisms underlying these phenomena. By investigating these regions further, future studies can shed light on the broader network dynamics associated with processing socially relevant information and elucidate their functional implications.

The unexpected finding that stronger social relationships were associated with more normative trait endorsements is intriguing and may be related to the results observed in the PCC and precuneus. The consistency between the endorsement findings and the PCC results suggests that there may be a common underlying mechanism driving both the behavioral and neural patterns. The PCC has been implicated in various aspects of social cognition, including self-referential processing, mentalizing, and integrating social information (Schilbach et al., 2008). It is possible that the PCC plays a role in aligning one's perceptions and behavioral responses with social norms and expectations, particularly in the context of close relationships. The normative patterns observed in both the endorsement data and the PCC neural representations may reflect this social alignment process. Furthermore, this speaks to the conflicting results that have been

observed in behavioral studies investigating normative trait profiles, with some of these investigations showing more normative trait assignment over time and others showing less. (Biesanz et al., 2007; Human et al., 2020). These findings highlight the interplay between explicit behavioral responses and implicit neural representations in social cognition. Future studies should investigate the relationship between these different levels of analysis, particularly focusing on the role of the PCC in integrating social information and shaping both behavioral and neural patterns in the context of close relationships.

While this study provides evidence for the influence of social relationship strength on normative social cognition, it is also important to acknowledge a limitation in our design. All participants within each group possessed some level of prior acquaintance with all group members. As a result, there were no true strangers in our paradigm, which limits our ability to compare our findings to that of representations seen in the impression formation literature (Brooks & Freeman, 2019; Schiller et al., 2009). Nevertheless, our findings demonstrate that despite this lack of complete unfamiliarity, the distinctiveness of representations still corresponded with the overall strength of social relationships. We demonstrated that there is indeed significant variability observed in the overall strength of relationships between perceivers and targets. By intentionally recruiting diverse groups, including individuals from different contexts such as friend groups, student groups, and work groups, we aimed to capture a range of relationship strengths. While this variability provides a more ecologically valid representation of real-world relationships, it can also complicate the interpretation of findings and may limit the generalizability of our results to specific relationship types. It is crucial to consider these limitations when interpreting the implications of our study and future research should aim to

address these issues for a more comprehensive understanding of the brain's representation of close others.

Summary

In summary, our research highlights a nuanced distinction in the neural mechanisms underlying social cognition, particularly the divergent roles of anterior and posterior regions in processing close others. Anterior regions, such as the vmPFC and ventral anterior insula, are shown to create unique, individualized representations of close others, emphasizing the brain's capacity for integrating affective and cognitive information to support deep personal connections. Conversely, posterior regions, including the PCC and precuneus, demonstrate a tendency towards more generalized processing with stronger social ties, suggesting a complementary but distinct cognitive function from the anterior regions. Interestingly, our behavioral findings reveal that stronger social relationships are associated with more normative trait endorsements, which appears to align with the neural patterns observed in the posterior regions. This divergence between the behavioral and anterior neural results underscores the inherent interplay between explicit responses and implicit neural representations in social cognition. The distinction between anterior and posterior brain regions challenges previous assumptions that these areas operate uniformly to support social cognition. Instead, our findings propose that anterior and posterior brain regions engage in different types of cognitive processing when representing close others, indicating a more intricate and differentiated neural basis for person perception and social cognition more broadly. This work may inform future studies on the neural basis of strategies that are employed as social relationships evolve from zero acquaintance to intimate bonds, and highlights the importance of considering both

behavioral and neural measures to gain a comprehensive understanding of social cognitive processes.

Chapter IV
Decoding Person Identity of Known Others

This work was co-authored with Aussie D. Frost & Robert S. Chavez. I was the lead author of the publication and responsible for the writing of the manuscript with edits provided by the co-authors. I led the methods and analyses with input and assistance from the co-authors.

Introduction

The human brain has evolved an extraordinary capacity to represent critical information in its environment and use these representations to facilitate goal-directed behavior (Cisek, 2019; Eichenbaum, 2004; Pessoa & Adolphs, 2010; Ranganath & Ritchey, 2012; Robinson et al., 2022). As a highly social species, our brains have had to adapt to significant social pressures which necessitates the development of representational mechanisms to distinguish between known social acquaintances (Cheng et al., 2023; Cikara et al., 2017; Courtney & Meyer, 2020; Kumaran et al., 2016; Thornton et al., 2019; Thornton & Mitchell, 2017). While there have been efforts to decode the neural representations underlying person identity, much of this work has focused solely on face or voice recognition (Anzellotti et al., 2014; Anzellotti & Caramazza, 2017; Axelrod & Yovel, 2015; Carlin, 2015; Tsantani et al., 2019). However, a person's identity extends far beyond their appearance or voice. Our perceptions of others are rich with abstract detail that captures the whole range of human emotion and experience. This implies a potential to decode a more internalized holistic representational space of our subjective reflections of others, which encompasses the full spectrum of abstract person identity and likely involves distinct neural structures from those implicated in external face and voice recognition alone.

The development of MVPA methods have advanced our understanding of the neural representations that underlie the cognitive processes that they serve (Haxby, 2012; Haxby et al., 2001; Kamitani & Tong, 2005; Norman et al., 2006). Whereas traditional univariate methods are instrumental in identifying the involvement of specific brain regions in mental states or cognitive

processes, MVPA allows for the investigation of spatial patterns within these regions that dynamically constitute the processes themselves. For instance, a univariate investigation might show that the mPFC is, on average, more active when contemplating social actors as opposed to everyday objects (Harris et al., 2007). However, pattern classification techniques allow us to peer into the mPFC itself and test whether this brain region employs distinct spatial patterns of activity to differentiate stimuli within the social actors category, despite each stimulus evoking similar overall average activity. This approach introduces a nuanced layer to the interpretation of a brain region's role, suggesting that the computations in this neural space are likely integral to representing or distinguishing one person from another.

Much of the research using MVPA to investigate the spatial patterns that are indicative of person identity has focused primarily on decoding face and voice stimuli (Anzellotti et al., 2014; Anzellotti & Caramazza, 2017; Axelrod & Yovel, 2015; Carlin, 2015; Tsantani et al., 2019). Many of these studies have employed a localization approach, first identifying the regions that exhibit maximal activity in response to face or voice stimuli and then conducting the classification analyses within these localized areas. These studies have demonstrated the ability to decode stimulus identity within regions such as the fusiform face area and the superior temporal sulcus (Anzellotti et al., 2014; Axelrod & Yovel, 2015). Efforts to uncover modality-invariant representations of person identity that are consistent across both visual and auditory stimuli have also highlighted the anterior temporal lobe as a potential integrative hub (Anzellotti & Caramazza, 2017; Tsantani et al., 2019). These studies have provided valuable insights into the neural mechanisms underlying person identity processing but they have primarily relied on isolated face or voice stimuli which may not fully capture the richness and complexity of person representations in real-world social contexts. The abstract, holistic nature of person identity,

which encompasses a wide range of internalized subjective reflections and experiences, likely involves neural structures and processes that extend beyond those identified through face and voice recognition alone.

The current study seeks to go beyond the limitations of previous research by investigating the neural representations of real, personally known individuals, evoked by abstract considerations. Employing a novel round-robin design allowed us to use stimuli that represented a range of real people that were personally known to each subject (Chavez & Wagner, 2020; Guthrie et al., 2022), moving away from the traditional reliance on isolated face or voice stimuli. Furthermore, the trait adjective task used to elicit cognitive processes involved in person perception that was used in this study (Kelley et al., 2002), demands a more nuanced consideration of the target "other" than simple face or voice recognition. It requires the subject to engage with their subjective reflections, memories, and experiences as they determine whether a specific trait word accurately describes the target individual. This approach enables a unique investigation into the creation and storage of abstract representations of known others in the brain, as well as the reliability of decoding these spatial patterns in neural regions beyond those that have been identified in similar previous work (Chavez & Wagner, 2020; Guthrie et al., 2022; Wagner et al., 2019).

We hypothesize that neural regions consistently shown to be involved in abstract social cognition will contain spatial patterns of brain activity that are reliable enough to decode the identity of personally known others (Amodio & Frith, 2006; Heatherton et al., 2006; Mitchell et al., 2002; R. Saxe & Kanwisher, 2003; Van Overwalle, 2009). This includes key nodes of the default mode network, such as the mPFC, TPJ, PCC, and the precuneus. These regions have been associated with various types of social cognition and person perception, such as mentalizing

(Saxe & Kanwisher, 2003; Saxe et al., 2009; Schurz et al., 2021), perspective-taking (Dumontheil et al., 2010; Jääskeläinen & Kosonogov, 2023; Santiesteban et al., 2012), trait inference (Ma, Baetens, Vandekerckhove, Kestemont, et al., 2014; Marquine et al., 2016; Raykov et al., 2021), and the processing of social knowledge (Adolphs, 2009; Arioli et al., 2021; Freeman et al., 2010). Furthermore, we hypothesize that the neural representations of person identity will be reliable and robust, regardless of whether a specific trait is endorsed or not for a particular individual on any given trial. This is based on the prediction that the abstract, holistic representation of a person's identity is likely to be activated whenever that person is thought about, irrespective of the trait being endorsed.

In addition to investigating the overall decoding accuracy and the impact of trait endorsement on the neural representations of personally familiar others, we also aim to explore the potential influence of social relationship strength on the classifier's performance. Previous research has shown that the strength of social connections can modulate various aspects of social cognition (Guthrie et al., 2022; Krienen et al., 2010; Parkinson et al., 2018). We hypothesize that the strength of the social relationship between the perceiver and each target individual will increase the sensitivity of the classifier in correctly identifying specific targets, particularly in brain regions involved in social cognition.

To test these hypotheses, we will employ a whole-brain searchlight procedure, which will allow us to identify the parts of the brain supporting locally distributed neural representations of person identity in an unbiased, data-driven manner (Kriegeskorte et al., 2006). Rather than relying on a priori regions of interest, this approach will enable us to identify brain regions containing reliable spatial patterns of activity associated with specific personally known others,

without any prior assumptions about their location. Within each searchlight object, we will use a support vector machine (SVM) classifier to predict the identity of the individual being thought about based on the spatial pattern of brain activity (Kriegeskorte et al., 2007). By iterating the searchlight object across the entire brain, we will create a map of decoding accuracies, revealing the regions that contain the most reliable person identity representations.

To investigate the robustness of these representations and the potential influences of trait endorsement, we will conduct separate analyses for endorsed and non-endorsed trials using the same searchlight procedure and also employ the same type of classification on trial-level models. Additionally, by utilizing the trial-level searchlight procedure, we aim to uncover the potential role of social relationship strength in shaping the classifier's sensitivity in identifying personally familiar others. This comprehensive approach aims to provide a rigorous investigation of the neural representations underlying the abstract concept of person identity, shedding light on the brain regions involved in encoding personally known others and elucidating the nature and stability of these representations across different cognitive and social contexts.

Methods

Participants

The same sample of participants described in Chapter 3 were utilized for this portion of the study. Due to the nature of this analysis however, the 2 subjects that were excluded in chapter 3 because of unusable behavioral data were included in the portions of this analysis that did not include the utilization of behavioral data. This left a final N = 113 subjects in total with at least five participants per group for the main decoding analysis.

Procedures

Behavioral and Neuroimaging

The experimental procedures for the behavioral and neuroimaging sessions are identical to those described in Chapter 3.

Analysis I – Classification Accuracy

Preprocessing

Functional imaging data acquired for the fMRI task underwent preprocessing and estimation of voxel responses using FSL (S. M. Smith et al., 2004). The preprocessing steps included mean-based intensity normalization, high-pass filtering (Gaussian-weighted least-squares straight line fitting with sigma = 100 s) and spatial smoothing with a 4-mm full-width at half-maximum (FWHM) Gaussian kernel.

Preprocessing also involved correcting for spatial distortions using field map unwarping and aligning the functional data to each participant's anatomical scan. This alignment was achieved through boundary-based registration (Greve & Fischl, 2009) in conjunction with a linear registration using FSL's FLIRT tool.

First-Level Analysis

First-level task-specific analyses were conducted in each participant's native space to prepare the data for subsequent searchlight classification analyses. In these analyses, parameter estimates were independently calculated for each of the five group members within each of the five runs. By estimating the neural response to each target person separately in each run, we obtained a set of five parameter estimates for each voxel and each run. These parameter estimates captured the unique neural patterns associated with thinking about each specific group member, allowing us to investigate whether these patterns could be used to decode the identity of

the target person in a given trial. All estimates from runs 2-5 were then registered via a linear transformation into the same native space as run 1 to ensure that all runs were functionally aligned for the searchlight analysis.

Overall Classification Accuracy Analysis

Searchlight and SVM Classification

To identify brain regions that could reliably decode the identity of the target person being thought about during the fMRI task, we performed a searchlight analysis using the BrainIAK library (Kumar et al., 2021). The searchlight technique involved moving a small cubic "searchlight" across the brain, extracting local patterns of brain activity, and testing how well these patterns could distinguish between the different target individuals.

The searchlight cube extended 1 voxel in each direction away from the center voxel, resulting in a cube size of 3 voxels x 3 voxels x 3 voxels (27 voxels in total). Given the voxel size of 2 mm, the physical dimensions of the searchlight cube were 6 mm x 6 mm x 6 mm. This size was chosen to capture local patterns of brain activity while maintaining a reasonable level of spatial specificity. The searchlight analysis was performed in each participant's native brain space to preserve individual differences in brain anatomy and function and the searchlight cube was moved across a native whole-brain mask, ensuring that all brain regions were considered in the analysis. At each location, the voxels within the searchlight cube were extracted, and their parameter estimates for each target person in each run were used as input features for a machine learning classifier.

We employed a linear Support Vector Machine (SVM) classifier to determine how well the local patterns of brain activity could distinguish between the different target individuals. The

classifier was trained and tested using a leave-one-run-out cross-validation scheme, where the data from four runs were used for training and the remaining run was used for testing. This process was repeated five times, with each run serving as the test set once. The mean classification accuracy across all cross-validation folds was calculated for each searchlight position, providing a measure of how well that local region could decode the identity of the target person.

To ensure the reliability of the classification results, we set a minimum threshold for the number of voxels required within the searchlight cube. If the number of voxels in the searchlight cube at a particular location was below this threshold (set to 14 voxels in this study), the classification accuracy was not calculated for that position. This step helped to avoid spurious results arising from searchlight cubes that extended beyond the edge of the brain or included too few voxels to provide meaningful information.

The searchlight analysis yielded individual searchlight maps for each participant, where each voxel value represented the mean classification accuracy of the local region centered around that voxel. These individual searchlight maps were then subjected to group-level analysis to identify consistent patterns of target identity decoding across participants.

Group-level Analysis

To assess the consistency of the searchlight accuracy maps across participants, we performed a group-level analysis using nonparametric permutation testing. First, all individual accuracy maps were transformed from native space to the standard MNI152 space, ensuring that the maps were aligned to a common spatial template for group-level comparisons. The transformation parameters used for this step were derived from the registration of each

participant's functional data to their anatomical data and then to the standard space during the preprocessing stage.

The transformed accuracy maps were then combined into a single 4D dataset, where each 3D volume represented a participant's accuracy map in standard space. To facilitate group-level comparisons, we normalized the accuracy values by subtracting the chance accuracy level (0.20) from each voxel, centering the values around zero. Positive values indicated above-chance accuracy, while negative values indicated below-chance accuracy. Nonparametric permutation testing was performed using the randomise tool in FSL to determine the statistical significance of the group-level accuracy maps. A one-sample t-test was used to compare the group-level accuracy values to zero (chance level) at each voxel. Uncorrected p-values were initially computed, and a cluster-forming threshold of $t = 5.24$ (corresponding to an uncorrected p-value of 0.005; $df = 112$) was applied to identify clusters of voxels with significant above-chance accuracy.

A binary brain mask derived from the standard MNI152 template was used to restrict the analysis to voxels within the brain, excluding non-brain regions from the statistical inference. This mask was slightly dilated to ensure adequate coverage of the brain. Permutation testing was carried out with 5000 permutations to build a null distribution of the test statistic. In each permutation, the participant labels were randomly shuffled, and the test statistic (t-value) was computed. To correct for multiple comparisons and identify significant clusters of above-chance decoding accuracy at the group level, we employed the cluster correction option implemented in FSL. This approach uses a cluster-based threshold to determine significant regions while controlling for the family-wise error (FWE) rate. The final statistical maps were thresholded at

an FWE-corrected cluster significance level of $p < 0.01$, identifying brain regions with significantly above-chance decoding accuracy at the group level.

Endorsement-Specific Classification Accuracy Analysis

To investigate the effect of trait endorsement on the decoding of target individuals, we performed an additional searchlight analysis that separately considered trials where subjects endorsed a trait for a target and trials where subjects did not endorse the trait. Subjects who either endorsed all traits or no traits for a given target in at least one run were excluded from this analysis, as it would be impossible to model the contrast between endorsed and non-endorsed trials in these cases. Out of the 113 subjects included in the overall accuracy analysis, 91 subjects remained for the endorsement-specific analysis after applying this exclusion criterion.

For the remaining subjects, we created independent estimates for each target, separately for endorsed and non-endorsed trials, yielding five estimates for each condition per run. The first-level analysis followed the same procedure as the overall accuracy analysis but with the additional separation of endorsed and non-endorsed trials. The searchlight and SVM classification were then performed separately for endorsed and non-endorsed traits, using the same parameters and cross-validation scheme as in the overall accuracy analysis. This allowed us to compare the decoding accuracies between the two conditions and assess the impact of trait endorsement on target decoding.

Group-level analyses were conducted independently for endorsed and non-endorsed conditions using nonparametric permutation testing. Individual accuracy maps for each condition were transformed to standard MNI152 space, combined into separate 4D datasets, and normalized by subtracting the chance accuracy level. Permutation testing with 5000 permutations

and a cluster-forming threshold of t = 5.26 (uncorrected p = 0.005; df = 90) were applied to each condition. The resulting statistical maps identified brain regions with significantly above-chance decoding accuracy for endorsed and non-endorsed trials separately. Comparing these maps allowed us to determine whether trait endorsement modulated the decoding of target individuals and to identify brain regions with differential decoding accuracy between the two conditions.

Results

We conducted three main analyses to investigate the neural decoding of social information about personally familiar others. First, we examined the overall decoding accuracy for identifying specific target individuals based on brain activity patterns. Second, we investigated the effects of trait endorsement on decoding accuracy by separately analyzing trials with endorsed and non-endorsed traits. Finally, we assessed the convergence between the endorsed and non-endorsed trait decoding results. It is important to note that all group-level analyses were conducted with a FWE cluster correction using a threshold at a t-critical value corresponding to an alpha of 0.005. Furthermore, the brain regions displayed in the figures (Figures 11-13) represent only those clusters that survived a statistical threshold of $p < .01$ after multiple comparison corrections were applied.

Overall Decoding Accuracy

The searchlight analysis revealed significant above-chance decoding accuracy for target individuals in several brain regions (Figure 13). Notably, key nodes of the default mode network exhibited consistent involvement in the decoding of person identity. The mPFC, PCC, and TPJ all showed significant decoding accuracies, with average accuracies ranging from 0.215 to 0.300 across subjects.

Endorsed Trait Decoding

When considering only trials with endorsed traits, we found significant above-chance decoding accuracy in similar regions as the overall analysis (Figure 14). Regions of the default mode network, including the mPFC, PCC, and TPJ, again demonstrated significant decoding accuracies, ranging from 0.215 to 0.266. This indicates that the neural representation of personally familiar others is robust even when focusing solely on traits that participants endorsed as characteristic of the target individuals.

Non-Endorsed Trait Decoding

The searchlight analysis of trials with non-endorsed traits also yielded significant above-chance decoding accuracy (Figure 15). The default mode network regions, particularly the mPFC, PCC, and TPJ, again showed significant decoding accuracies that ranged from 0.215 to 0.270. These results suggest that the neural patterns associated with target individuals are distinguishable even when considering traits that participants did not endorse as characteristic of those individuals.

Endorsement Convergence

To assess the convergence between the endorsed and non-endorsed trait decoding results, we examined the overlap of significant regions (Figure 16). Specific portions of the default mode network, including the mPFC and PCC, exhibited consistent involvement across both endorsement conditions. This convergence highlights the central role of the default mode network in representing personally familiar others, regardless of whether the traits being considered are endorsed or not.

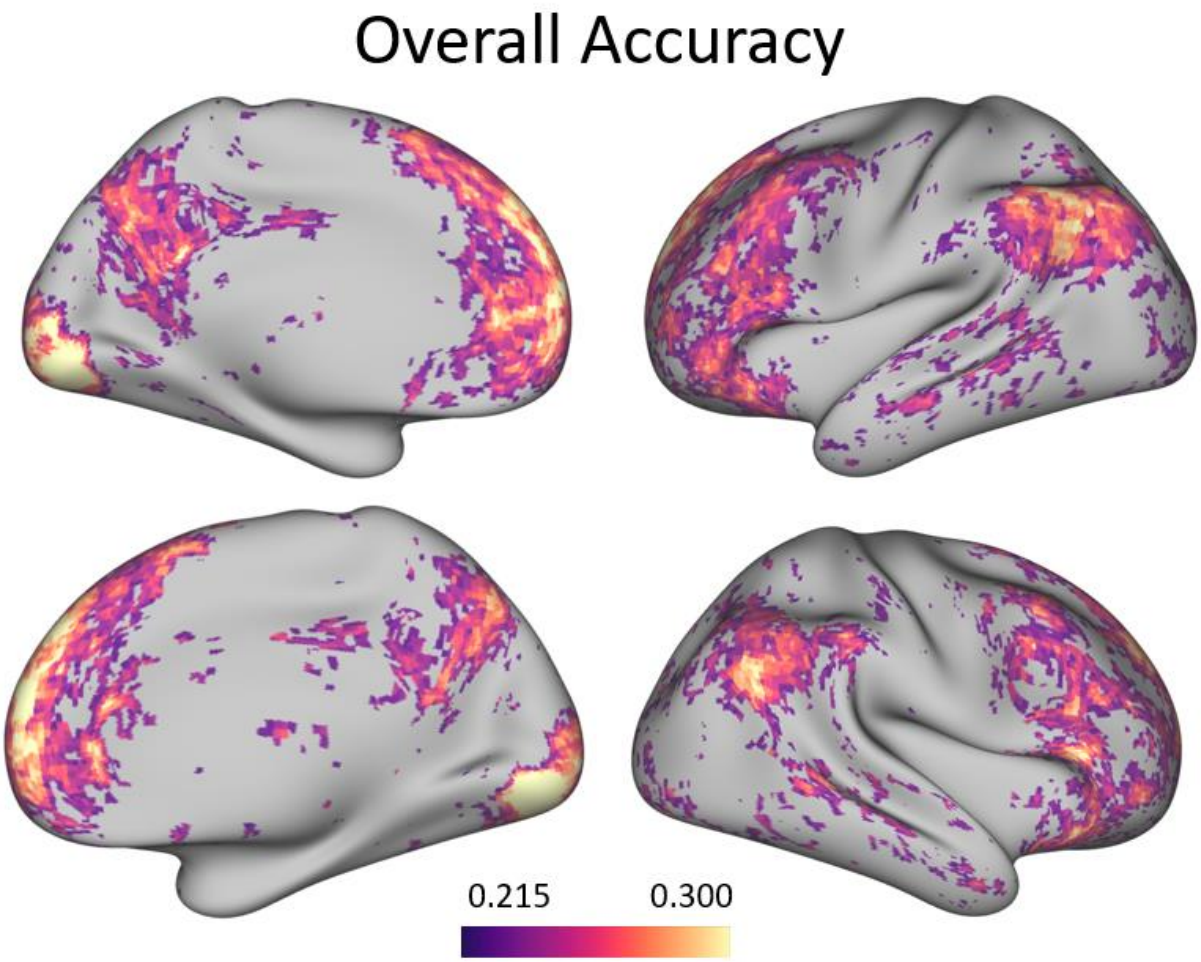

Figure 13. Overall Accuracy - Brain regions showing significantly above-chance decoding accuracy for target individuals based on the searchlight analysis of run estimates. The color scale represents the average decoding accuracy across subjects within significant clusters ($p < 0.01$, corrected for multiple comparisons).

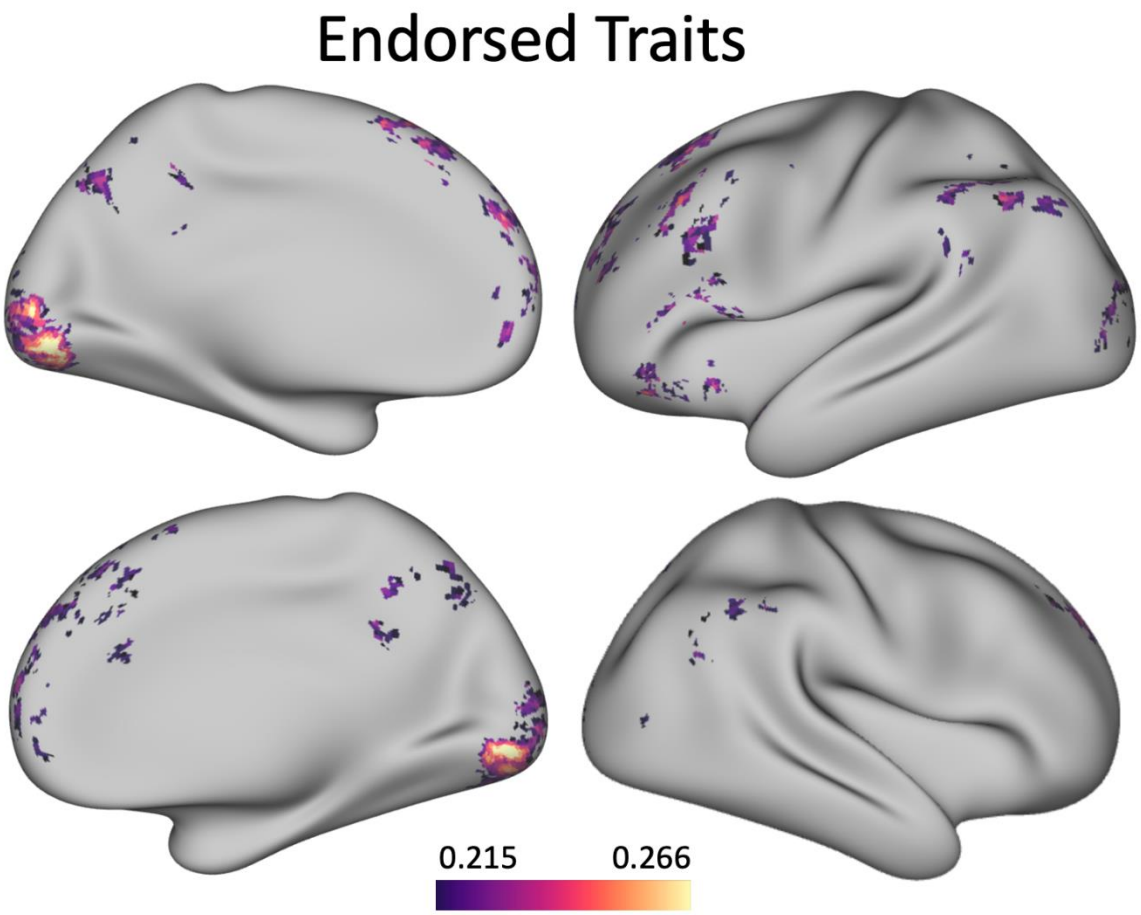

Figure 14. Endorsed Traits - Brain regions showing significantly above-chance decoding accuracy for target individuals based on the searchlight analysis of trials with endorsed traits. The color scale represents the average decoding accuracy across subjects within significant clusters ($p < 0.01$, corrected for multiple comparisons).

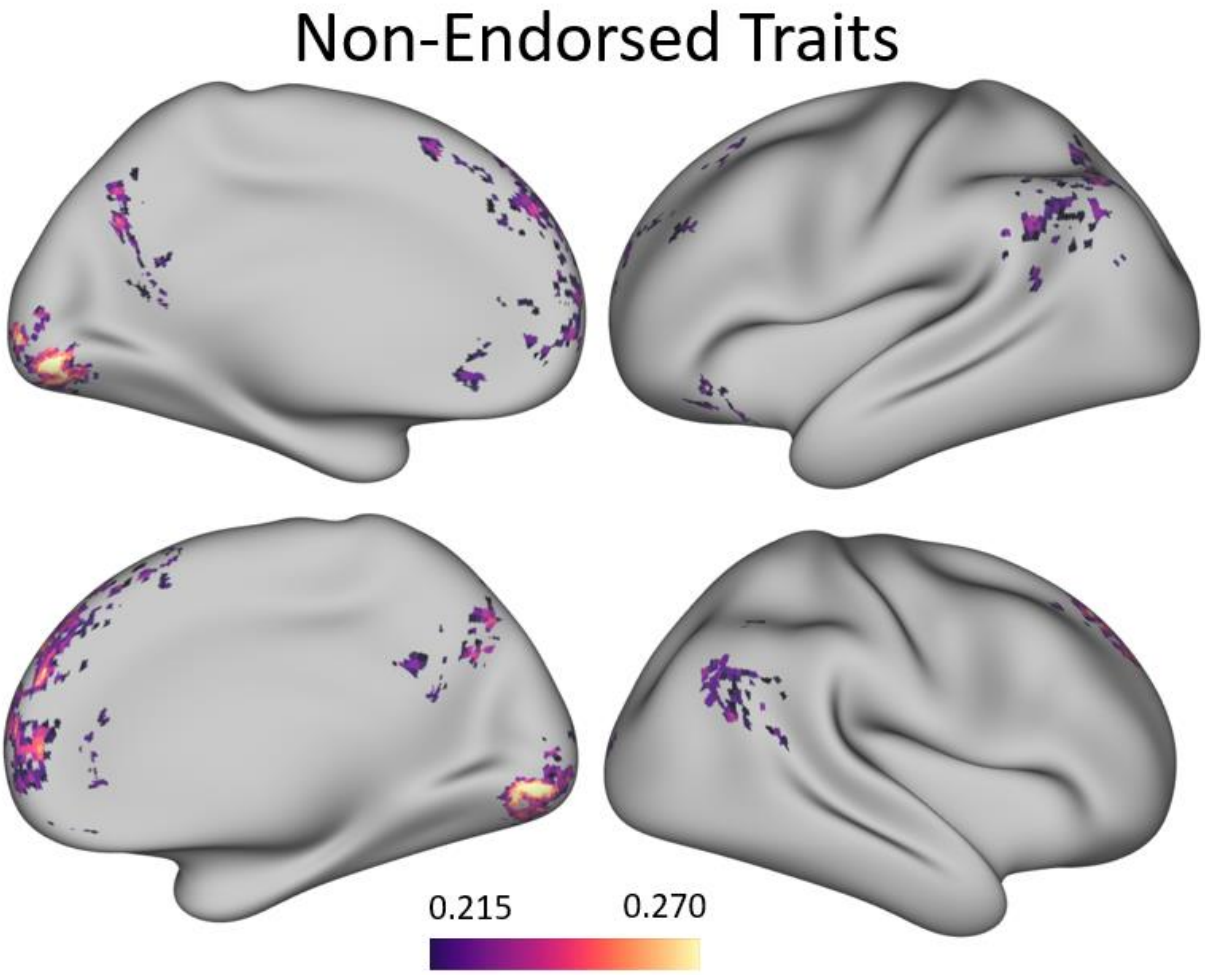

Figure 15. Non-Endorsed Traits - Brain regions showing significantly above-chance decoding accuracy for target individuals based on the searchlight analysis of trials with non-endorsed traits. The color scale represents the average decoding accuracy across subjects within significant clusters ($p < 0.01$, corrected for multiple comparisons).

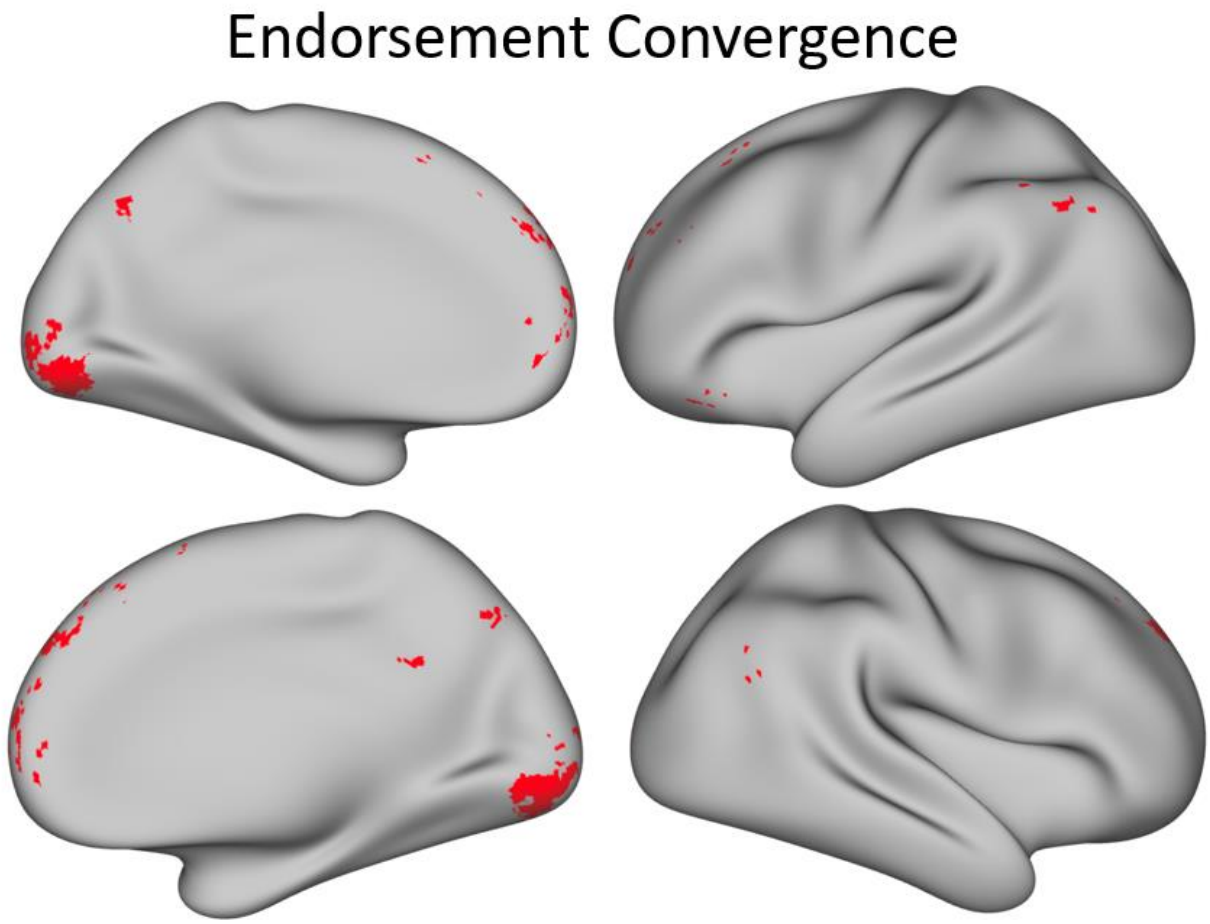

Figure 16. Endorsement Convergence - Brain regions in which the significant decoding accuracy for endorsed and non-endorsed traits overlapped.

Analysis II – Social Relationship Strength

To investigate whether individual differences in social relationship strength between the perceiver and target could account for variability in the overall accuracy of the classifier, we utilized the sensitivity metric. Sensitivity, calculated per target as the ratio of true positives to true positives plus false negatives, provides insight into the underlying factors driving the classifier's accuracy. A sensitivity score of 1 indicates that the classifier correctly identified a

target as a true positive every time, with no false negatives reported for that target. However, if the classifier misclassifies that target or confuses it with another target, the sensitivity metric would move closer to zero. We hypothesized that higher social relationship strength would result in higher sensitivity measures for a particular target, meaning that it would be correctly identified when it was a true positive and not be confused with other targets.

First Level Analysis

Our initial analysis, which used a single estimate per target per run, limited the assessment of classifier sensitivity. While this approach allowed us to train and test on run-level estimates, which ideally have a higher signal-to-noise ratio, the hold-out run used for testing only contained one instance of each target. This resulted in the sensitivity metric being essentially identical to accuracy because the target was either guessed correctly or not. To overcome this limitation and gain a more comprehensive understanding of the factors influencing the decoding of personally familiar others, we employed a trial-level modeling approach.

By modeling each trial as a separate contrast above baseline in FSL, we obtained 12 estimates per target per run, accounting for all other trials in the modeling procedure. This trial-level modeling approach allowed us to test the reliability of the classifier accuracy on estimates that presumably have lower signal-to-noise ratios compared to run estimates and to investigate the potential influence of social relationship on the decoding accuracy. Testing on 60 estimates per target and using a hold-out testing sample with 12 estimates of each target enabled a more reliable metric of sensitivity, as it allowed for the consideration of both true positives and false negatives. This finer-grained approach provided a more robust assessment of classifier performance, allowing us to investigate the role of social relationships in modulating the overall accuracy of the classifier.

SVM Classifier Performance

The procedure for employing the SVM classifier with the searchlight approach in this analysis was identical to the procedure used in the previous analyses, with the only difference being the size of the training and testing sets. As in the earlier analyses, the classifier was trained and tested using a leave-one-run-out cross-validation scheme, where the data from four runs were used for training and the remaining run was used for testing. This process was repeated five times, with each run serving as the test set once.

Group-level analyses to test which regions were significant above chance were also conducted using the same procedures as in the previous analyses. Nonparametric permutation testing was performed using the randomise tool in FSL, with a one-sample t-test comparing the group-level accuracy values to chance level at each voxel. The same cluster-forming threshold and multiple comparison correction methods were applied to identify significant clusters of above-chance decoding accuracy.

The key difference in this analysis was the inclusion of sensitivity measures, which were retrieved from the classifier modeling procedure each time it was run, averaged across the five cross-validation folds. The sensitivity value was assigned to the center voxel of the searchlight object, resulting in one sensitivity metric for every voxel for every target. This yielded a 3D NIFTI file for each target, representing the average sensitivity for that target at each location in the brain. These sensitivity measures provided additional insights into the classifier's performance and allowed for the investigation of the relationship between social relationship strength and the classifier’s ability to correctly identify specific targets.

Cluster-Based ROI Approach

To test our hypothesis about the relationship between social relationship strength and the classifier's sensitivity, we focused on brain regions that passed the group-level significance testing for decoding accuracy. The group-level analysis, as described in the previous section, allowed us to determine which regions were significant above chance. However, to test each region within the significant results independently, we needed to separate them into distinct clusters.

To achieve this, we calculated the mean accuracy for the SVM classifier across subjects and then applied a cluster algorithm from FSL to the resulting 3D nifti file. The cluster algorithm was applied with a threshold of 0.215 accuracy, which sufficiently separated the significant regions into three distinct ROIs: the mPFC, PCC, and primary visual cortex (V1). This approach allowed us to investigate each region independently and test the relationship between social relationship strength and classifier sensitivity in a more targeted manner. For each of these three clusters, we calculated the average sensitivity per target, which served as the dependent variable in subsequent analyses.

Social Relationship Prediction

To assess the impact of social relationship on the decoding accuracy of personally familiar others, we employed linear mixed-effects models with random intercepts using the lme4 package in R. The social relationship scores used as predictors in these models were calculated by averaging the social relationship strength scores across four metrics (knowing, friendship, liking, and similarity) that were reported by the perceiver for each of the five targets in the

behavioral session of the experiment. This resulted in a single social relationship score for each perceiver-target pair, representing the overall strength of their social bond.

Separate linear mixed-effects models with random intercepts were employed in each of the three clusters (mPFC, PCC, and V1) to test them independently. For each cluster, the average sensitivity per target was used as the dependent variable, while the social relationship scores served as predictors. Subjects were nested within groups as a mixed effect to account for the hierarchical structure of the data. This approach allowed us to investigate whether the strength of the social relationship between the perceiver and a particular target modulated the sensitivity of the classifier in decoding the target's identity based on neural activity patterns within each specific cluster.

Results

Analysis II – Social Relationship Strength

To investigate the influence of social relationship strength on the decoding accuracy of personally familiar others, we conducted an additional analysis using a trial-level modeling approach. This approach allowed us to assess the classifier's sensitivity in identifying specific targets and explore the potential impact of social relationship strength on decoding performance. First, we attempted to replicate the original decoding accuracy results from Analysis 1 using the trial-level data. Then, we assessed whether the social relationship strength between the perceiver and each target was predictive of the sensitivity measure per target produced by the classifier during every iteration. The sensitivity measure, which reflects the classifier's ability to correctly

identify a target when it is a true positive and also not confuse that target with other targets, provides insight into the factors influencing the decoding of personally familiar others.

Overall Accuracy with Trial-Level Modeling

The searchlight analysis using trial-level estimates revealed significant above-chance decoding accuracy for target individuals in regions similar to those identified in the original analysis that used run-level estimates (Figure 17). The default mode network, including the mPFC, PCC, and TPJ, exhibited consistent involvement in the decoding of person identity. However, the average decoding accuracies within these regions were slightly lower compared to the original analysis, ranging from 0.21 to 0.24 at the group level. This reduction in decoding accuracy is likely due to the noisier nature of the trial-level estimates compared to the run-level estimates used in the previous analysis. Despite this reduction, the replication of the main findings using trial-level modeling demonstrates the robustness of the neural representations of personally familiar others within the default mode network.

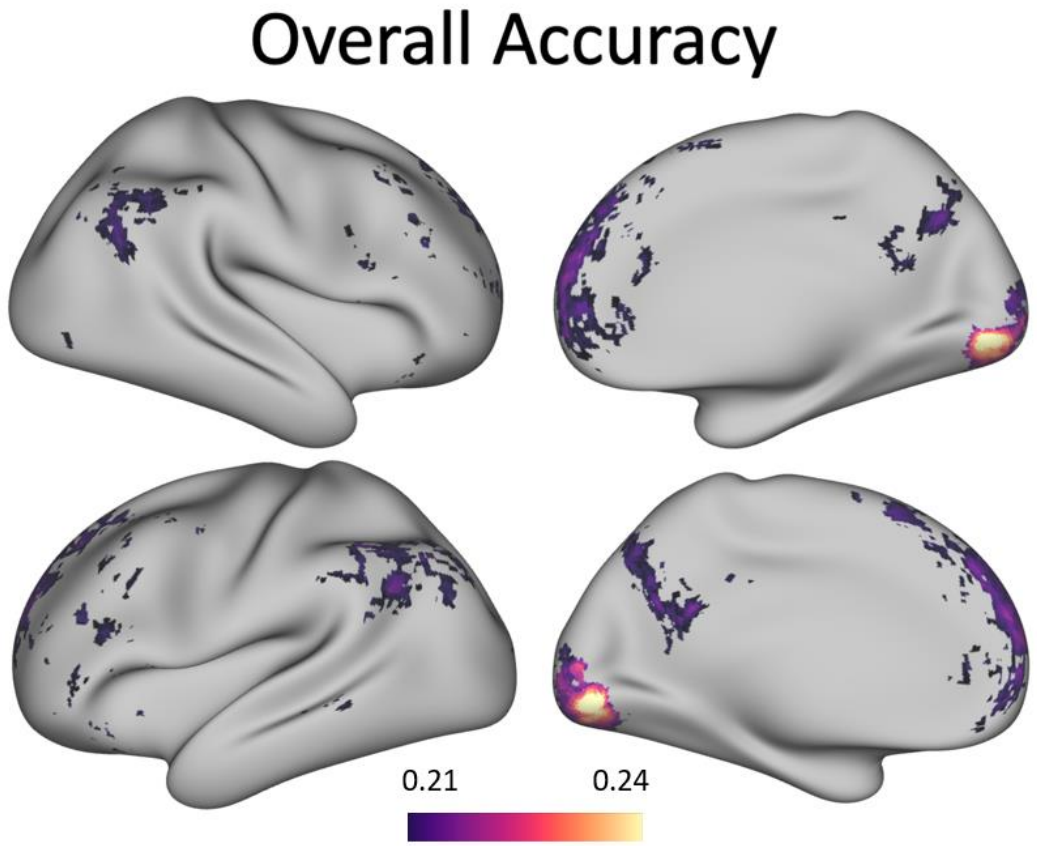

Figure 17. Overall Accuracy with Trial-Level Modeling - Brain regions showing significantly above-chance decoding accuracy for target individuals based on the searchlight analysis using trial-level estimates. The color scale represents the average decoding accuracy across subjects within significant clusters ($p < 0.01$, corrected for multiple comparisons).

Sensitivity and Social Relationship Strength

Next, we focused on three significant clusters identified in the group-level analysis to examine the relationship between social relationship strength and the classifier's sensitivity in identifying specific targets. These clusters included the mPFC, PCC, and V1 (Figure 18). Linear mixed-effects models with random intercepts were employed to test the impact of social relationship scores on the average sensitivity per target within each cluster.

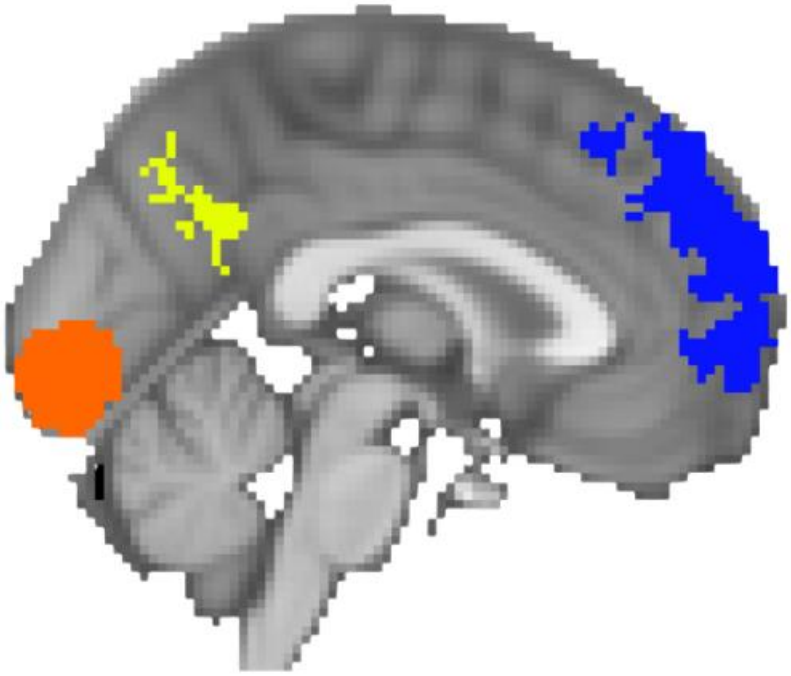

Figure 18. Cluster-Based ROIs for Sensitivity Analysis - This figure illustrates the three distinct clusters used for the sensitivity analysis, which were derived from the mean accuracy map using a cluster algorithm with a threshold of 0.215. The clusters include the mPFC; blue, PCC; yellow, and V1; orange. These ROIs were used to independently investigate the relationship between social relationship strength and the classifier's sensitivity in identifying specific targets within each region.

Unexpectedly, the results revealed a significant negative relationship between social relationship scores and classifier sensitivity in the PCC ($\beta = -0.004$, SE = 0.002, $p = 0.037$). This finding suggests that higher social relationship strength between the perceiver and a particular target was associated with lower sensitivity of the classifier in correctly identifying that target based on neural activity patterns within the PCC. Alternatively, this also states that the classifier was most sensitive in this region when social relationship was low. No significant relationships were observed between social relationship scores and classifier sensitivity in the mPFC or V1.

Discussion

As we navigate the social world, our brains must constantly process and represent information about the people around us. These representations likely encompass not only physical attributes such as facial features or voices but also the abstract aspects of a person's identity, including their traits, behaviors, and the subjective experiences we associate with them. In this study, we employed a round-robin fMRI design to investigate the neural representations of personally familiar others, aiming to identify the brain regions involved in the encoding of abstract representations of person identity. Our main finding demonstrates that regions known to be involved in social cognition, particularly key nodes of the default mode network, contain reliable spatial patterns of brain activity that can successfully decode the identity of specific individuals above chance. This result highlights the central role of these regions in representing the rich, multifaceted nature of person identity, evoked by subjective reflections of personally familiar others.

The robustness of our findings is underscored by the consistency of the decoding results across different analyses and conditions. First, we demonstrated that the neural representations of person identity remain stable and reliable, regardless of whether specific traits are endorsed or not for a particular target individual on any given trial. Separate searchlight analyses for endorsed and non-endorsed traits revealed significant above-chance decoding accuracies in the same key regions of the default mode network, including the mPFC, PCC, and TPJ. Moreover, we observed a substantial degree of overlap between the brain regions that exhibited successful decoding for both endorsed and non-endorsed traits. This suggests that the abstract, holistic representation of a person's identity is activated whenever that person is thought about, irrespective of the specific trait associations being considered. Second, we replicated our main

findings using trial-level estimates, which tend to be noisier than the run-level estimates used in the primary analysis. This further highlights the sensitivity of our methods in detecting reliable spatial patterns of brain activity associated with specific individuals, even in the presence of increased noise in the data.

The default mode network has been consistently implicated in various aspects of social cognition related to person perception, including mentalizing, perspective taking and the representation of social knowledge (Harris et al., 2007; Mars et al., 2012; Mitchell, Banaji, et al., 2005b; Nugiel & Beer, 2020; R. Saxe, 2006; Tamir & Mitchell, 2010; Thornton & Mitchell, 2018; Welborn & Lieberman, 2015). The mPFC, in particular, has been linked to the processing of person knowledge and the formation of impressions about others (Amodio & Frith, 2006; Dang et al., 2019; Heleven & Van Overwalle, 2019; Mitchell, Macrae, et al., 2006; Raykov et al., 2020, 2021). It has also been shown to be preferentially engaged when individuals make judgments about the traits, preferences, and mental states of others (Cloutier et al., 2011; Harris et al., 2005; Hassabis et al., 2014; Ma, Baetens, Vandekerckhove, Kestemont, et al., 2014; Schiller et al., 2009). Similarly, the PCC and TPJ have been implicated in various aspects of social cognition, including the retrieval of autobiographical memories, the representation of social contexts, and the understanding of others' perspectives (Carter & Huettel, 2013; Cavanna & Trimble, 2006; R. Saxe & Kanwisher, 2003; Schuwerk et al., 2017; Spreng et al., 2009). The successful decoding of person identity in these regions suggests that they contribute to the formation and retrieval of detailed, contextualized representations of personally familiar others, potentially integrating socially relevant information about the particular other.

It is important to address the unexpectedly high decoding accuracy observed in V1 as this result might suggest that stable internal visual representations of the targets are being captured in

V1. However, such an interpretation should be made with caution, as it is unlikely that V1 itself is encoding complex visual representations of individuals. If there were indeed stable internal visual representations of the targets evoked by the task, we would expect to find such representations further down the visual processing stream, in higher-order visual areas known to process more complex features and object categories (Grill-Spector & Weiner, 2014; Kanwisher & Yovel, 2006). Instead, the high decoding accuracy in V1 is more likely driven by the specific nature of the task used in this study. During the experiment, subjects were asked to judge whether a trait word associated with the name of a particular target was representative of that person. Crucially, the same name, with consistent visual properties such as the orientation of lines and edges, was presented for all trials associated with a given target. Given the well-established role of V1 in processing low-level visual features such as edges, lines, and contrast (Hubel & Wiesel, 1968; Kamitani & Tong, 2005), it is plausible that the high decoding accuracy in this region reflects the consistent visual properties of the presented names across trials and runs, rather than the encoding of stable internal visual representations of the targets themselves.

We had initially hypothesized that targets with higher social relationship strength would drive the accuracy of the classifier, such that perceivers would have more stable and well-defined representations of the people they are closest to. However, our results revealed the opposite pattern in the PCC, where lower social relationship scores were associated with higher classification accuracy. Consistent with the findings from chapter III, this finding suggests that, in the PCC, there may be a more consistent and generalizable pattern of neural activity across runs for individuals with whom we have weaker social relationships. Given the PCC's well-established role in memory retrieval and the integration of autobiographical information (Cavanna & Trimble, 2006; Leech & Sharp, 2013; Spreng et al., 2009; Svoboda et al., 2006), as

well as general person knowledge (Denny et al., 2012; Wagner et al., 2012), one possible interpretation is that, for personally familiar others with whom we have stronger social bonds, the retrieval of person knowledge may be more differentiated and specific to memories related to the traits being assessed in each run. In contrast, for acquaintances or individuals with whom we have weaker ties, the retrieval of person knowledge may be more general and consistent across runs, potentially relying on more stereotypical or categorical representations.

While this finding sheds light on one potential factor influencing the decoding accuracy in the PCC, it is important to note that it does not fully explain the decoding results observed in other key regions of the default mode network, such as the mPFC and TPJ. The absence of a significant relationship between social relationship strength and classifier sensitivity in these regions suggests that other factors, beyond the strength of social ties, may be driving the successful decoding of person identity in these areas. Further research is needed to investigate the specific aspects of person identity that contribute to the decoding accuracy in different default mode network regions and to explore the potential dissociation between the neural representations of personally familiar others based on the strength of social relationships.

While our study demonstrates significant above-chance decoding of person identity in key regions of the default mode network, it is important to acknowledge the limitations of these findings. Although we were able to decode the identity of personally familiar others with accuracy levels exceeding chance, it would be unreasonable to suggest that we can reliably predict who an individual is thinking about based solely on brain data under these experimental conditions. The highest decoding accuracy achieved in our study was approximately 0.30, which, while statistically significant, leaves considerable room for improvement in terms of predictive

power. However, it is noteworthy that the regions exhibiting above-chance decoding accuracy are consistent with those implicated in social cognition and person perception. This suggests that there is indeed some form of reliable person-specific social information encoded within this network that can be detected using pattern analysis techniques. (Haxby, 2012; Kriegeskorte et al., 2006) The fact that we were able to decode internally generated representations of personally familiar others with some degree of reliability, in the absence of explicit visual or auditory cues, offers new insights into the neural basis of social cognition and person representation, paving the way for future research in this domain.

To further advance our understanding of how the brain represents and processes person identity, future research should focus on refining the methods used in this study and exploring the specific types of internally generated social information that contribute to increased classifier accuracy. This may involve investigating the role of different cognitive processes, such as autobiographical memory retrieval, impression formation, and social reasoning, in shaping the neural representations of personally familiar others. Additionally, future studies could examine how factors such as the duration and quality of social relationships, as well as individual differences in social cognitive abilities, influence the stability and distinctiveness of person identity representations in the brain.

Summary

In conclusion, our study provides evidence for the successful decoding of internally generated representations of personally familiar others from patterns of brain activity, particularly within key regions of the default mode network. The robustness of our findings is underscored by the consistency of the decoding results across different analyses, including those

examining the effects of trait endorsement and the replication of the primary results with the use of trial-level models. Notably, the successful decoding of person identity was observed regardless of whether a specific trait was endorsed or not for a given target on a particular trial. This suggests that the neural representations of personally familiar others are stable and persistent, reflecting the inherent nature of person identity rather than momentary associations or evaluations. While our study reveals the unexpected finding that social relationship strength is predictive of classifier sensitivity in the PCC, with lower social relationship scores associated with higher decoding accuracy, this result highlights the multifaceted nature of person identity representations in the brain. Future research should build upon these findings by investigating the specific aspects of person identity that drive the decoding accuracy in different default mode network regions, exploring the relationship between neural representations and real-world social behavior, and examining the dynamics of person identity representations over time.

Chapter V

General Discussion

The self is a multifaceted cognitive construct shaped by our experiences and social interactions. As such, it serves as a powerful force in shaping the allocation of cognitive resources and the processing of self-related information (Cherry, 1953; Heatherton, 2011; H. Markus, 1977; Northoff et al., 2006; T. B. Rogers et al., 1977; Sui et al., 2013). The overarching goal of this dissertation was to investigate the neural mechanisms underlying the self's ability to influence cognitive processing and to explore how the brain differentially processes information related to the self and close or distant others. By leveraging advances in neuroimaging analysis techniques, computational methods, and novel experimental designs, we sought to uncover the computational role of the self and examine how self-related processing influences various cognitive processing streams. The studies presented in this dissertation provide evidence for the self as a unique and powerful cognitive construct that facilitates the recruitment of diverse cognitive networks to deeply process and modify self-relevant information, ultimately shaping our understanding of ourselves and others.

Chapter II demonstrated that self-relevant information significantly alters the way the brain processes information, leading to the deployment of more cognitive resources compared to unfamiliar information. We found widespread activation across the cortical hierarchy during self-related narratives, extending from low-level sensory processing to higher-order cognitive areas. Heightened activity was observed in regions associated with salience, attention, and the default mode network when participants listened to personally relevant narratives versus that of strangers or even dyadic partners. Increased neural synchrony was also found between dyad

members not only in key nodes of the default mode network, including the mPFC, PCC, and TPJ, but also in regions associated with attention and salience processing. This suggests that shared personal relevance, from the relationship with the dyadic partner, enhances the inter-subject synchronization of neural responses in brain areas involved in self-referential and social cognitive processes and also in regions responsible for attentional control and the detection of salient stimuli. These findings underscore the link between self-relevance and attention, highlighting the brain's ability to prioritize and deeply process information that holds personal significance. Moreover, the ISC and univariate overlap analyses suggest that these cognitive computations are generalizable across subjects.

Building upon these findings, chapter III, investigated how the strength of social relationships influences the creation of unique or normative representations of others in key regions known to be involved in social cognition. We observed that closer relationships resulted in more unique representations in the mPFC and anterior insula, areas associated with mentalizing and trait attribution (Chavez, 2021; Lau et al., 2020; Ma, Baetens, Vandekerckhove, Van der Cruyssen, et al., 2014b), indicating that these regions may be involved in the internalized process of individuation of highly self-relevant others. In contrast, more generalized representations emerged in posterior regions like the PCC, which interestingly aligned with our behavioral findings that revealed stronger social relationships were also associated with more externally indicated normative trait endorsements. These results suggest that while certain brain regions, such as the mPFC and anterior insula, may be involved in the internal individuation of close others, the PCC may be more involved in the generalized processes of conforming perceptions to cultural norms. This highlights the potential existence of differential processing streams in the brain when encoding information about others and emphasizes the ways in which

cognitive resources recruited for self-related stimulus processing may be utilized to modify pre-existing representational structures.

Finally, in chapter IV, we employed multivariate pattern analysis techniques to investigate the neural representations underlying the abstract concept of person identity. We found significant above-chance decoding accuracies in key regions of the default mode network, including the mPFC, PCC, and TPJ, when identifying specific target individuals based on brain activity patterns. These results were consistent across different analyses, indicating that the default mode network contains reliable information that can be used to differentiate personally familiar others from each other, even in the absence of visual or auditory cues. However, it is important to note that the accuracy levels for prediction were only slightly above chance, suggesting that while the information is reliable, it may not be as stable or persistent as initially thought. Interestingly, our sensitivity analysis revealed that lower social relationship scores were associated with higher decoding accuracy in the PCC, suggesting that the model may be more sensitive to correctly identifying others who are not close to the participant in this region. This potentially indicates that the model is able to differentiate between close and non-close others rather than distinguishing one person from another in this particular part of the default mode network. These results highlight the multifaceted nature of person identity representations in the brain and suggest that the strength of social relationships may modulate the consistency and distinctiveness of these representations in different default mode network regions.

The Self as a Powerful Cognitive Construct

The self has long been thought to be a powerful force in psychological reflection and social engagement. Philosophers and cognitive scientists provided solid theoretical frameworks that positioned the self as a central component in cognition, capable of marshalling the full battery of cognitive capabilities to serve it's purposes. (Beauvoir et al., 1989; James, 1890; Kant et al., 2000; Markus, 1977; Rogers et al., 1977). While modern neuroscience investigations have made significant progress in identifying the regions and networks involved in self-referential processing compared to those used for semantic understanding, (Kelley et al., 2002; Northoff et al., 2006; Wagner et al., 2012), there has been limited evidence demonstrating how these regions and networks work collectively to produce the behavioral changes observed with self-relevant material, especially when processing more naturalistic stimuli.

The studies presented in this dissertation aimed to connect the work that has been done, mostly in the social neuroscience domain regarding self-reference, with a more general cognitive framework. While the position of the self as a powerful construct has been intuitively understood by some for hundreds of years, a clear mechanistic account of how a self knowledge structure in the brain could influence attention or memory processes has yet to be fully understood. We provide evidence that shows the vastly different ways in which narrative content is processed when it is in direct reference to the self compared to when narratives are about close or unfamiliar others and demonstrated the involvement of not only regions known to be involved in self or social cognition but also ones involved in all of these other cognitive processes as well. The correlative work done here does not go so far as to reveal the exact mechanistic process by which this coordination across networks occurs, but it does highlight how pervasive this phenomenon is and how powerful self-related stimuli can be in altering the activity of the brain.

These results not only support the idea of the self as a central organizing principle in human cognition but also shed light on how the self may be guiding the construction of our experience from low level sensory processing to integrated high level narrative understanding through its control of attention and salience processing.

The univariate findings from chapter II provide evidence for the self as a powerful cognitive construct that influences the allocation of cognitive resources and the processing of self-related information. Our initial analysis contrasting each condition with the others revealed that self-narratives produced heightened activity compared to both partner and stranger narratives across the entire cortical hierarchy, from low-level sensory regions like the primary auditory cortex (A1) and even the thalamus to higher-order areas such as the default mode network. This widespread heightened activity was also observed when self and partner conditions were combined and contrasted with the stranger condition, albeit to a lesser extent than the self-only condition. These findings suggest that information relevant to the self, whether it pertains to oneself or a close other, modulates the signal-to-noise ratio, likely making the information more accessible for deeper processing. The involvement of diverse regions like A1, which primarily processes sound frequencies (Bendor & Wang, 2006; Formisano et al., 2008; Humphries et al., 2010), and the default mode network, which is involved in abstract representations and integration (Buckner & DiNicola, 2019; Hasson et al., 2015; Spalding et al., 2018; Vatansever et al., 2015), highlights the pervasive influence of self-relevance on information processing. Importantly, the heightened activity observed in regions known to be involved in attentional processing during self-relevant conditions may be key to understanding how the self exerts its influence on cognitive processing as a whole.

Our findings align with and extend existing cognitive neuroscience models and research on attention and self-referential processing. The Self-Attention Network model, proposed by Sui and Humphreys (2015), posits that the vmPFC, along with other regions such as the posterior superior temporal sulcus and the intraparietal sulcus, form a network that supports the rapid processing of self-relevant information. This model suggests that the social attention network plays a crucial role in the allocation of attentional resources to self-related stimuli and facilitates their preferential processing and integration with other cognitive processes (Sui & Humphreys, 2015a; Sui & Rotshtein, 2019). While the work done by Sui and colleagues provides a strong foundation for understanding the interplay between self-referential processing and attention, their studies employed highly controlled designs with simple stimuli. In contrast, our study utilized a rich, naturalistic design with emotionally charged information and personally meaningful content, more closely resembling the ways in which we engage with information in the real world. The heightened activation observed in the vmPFC, posterior superior temporal sulcus and the intraparietal sulcus regions during self-relevant narratives in our study provides support for the self-attention network model in a more ecologically valid context, demonstrating the engagement of this network in the processing of self-related information in a setting that better reflects real-life experiences.

Our findings of increased activity in sensory regions during self-relevant narratives are also consistent with previous cognitive neuroscience studies on attention. Research using functional neuroimaging has shown that attention can modulate activity in early sensory areas, such as V1 (Brefczynski & DeYoe, 1999; Hopfinger et al., 2000; Kastner et al., 1999; Luck et al., 1997; Moran & Desimone, 1985) and A1 (Fritz et al., 2007; Grady et al., 1997; Jäncke et al., 1999; Petkov et al., 2004; Schüller et al., 2023; Woldorff & Hillyard, 1991), leading to enhanced

processing of attended stimuli. For example, Hopfinger et al. (2000) demonstrated that covertly directing attention to one of two stimuli in the visual field resulted in increased activity in the corresponding retinotopic regions of the visual cortex for that stimulus compared to the activity produced by the identical stimulus in the unattended contralateral retinotopic location. Similarly, Petkov et al. (2004) found that attending to a particular auditory stimulus enhanced activity in the primary auditory cortex also suggesting that attention can modulate the processing of low-level sensory information. The heightened activation in sensory regions during self-relevant narratives in our study extends these findings to the domain of self-referential processing by showing how self-related information recruits attentional resources and enhances the processing of low-level sensory features.

The ISC results from our study provide further evidence for the self as a powerful cognitive construct by demonstrating that the processing of self-relevant information is consistent and generalizable across individuals. We found increased neural synchrony between dyad members in key regions of the default mode network that included the mPFC, PCC, and TPJ, when listening to personally relevant narratives compared to narratives from an unfamiliar individual. This heightened synchronization suggests that the brain's response to self-related information is not idiosyncratic but rather follows a common pattern across people. The consistent engagement of the default mode network across participants during self-relevant narratives indicates that this network plays a central role in the processing of self-related information in a way that may differ from the processing of non-relevant or unfamiliar social information (Buckner et al., 2008; Hasson et al., 2008; Raichle, 2015). The increased ISC in regions involved in sensory processing, such as A1 also suggests that the allocation of attentional resources to self-relevant information enhances the processing of low-level features in a

consistent manner across individuals (Ki et al., 2016; Regev et al., 2019). These findings highlight the robustness of the self as a cognitive construct and demonstrate that the neural mechanisms underlying the processing of self-related information are not only powerful but also generalizable across people.

Despite the evidence provided by the ISC results in the current study, it is important to acknowledge the inherent limitation of two-person neuroscience (Schilbach et al., 2013) approaches in isolating the self-condition. Due to the nature of ISC-like methods, the analysis always involves comparing one subject listening to themselves with another subject listening to them. This makes it challenging to directly examine the neural network architecture associated with processing purely self-related information. To circumvent this limitation, the current study included a close other condition, operationalizing the close other as self-relevant compared to a stranger. While this approach does not entirely isolate the self-condition, the robust areas of overlap observed between the ISC results and the univariate results, which were able to isolate the self, suggest that the self-relevant operationalization is likely a useful proxy for studying these processes. The consistency between the two methods indicates that the neural mechanisms underlying self-referential processing can be effectively investigated using the self-relevant operationalization, despite the inherent limitation of two-person neuroscience approaches.

To address the limitation of isolating the self-condition in the current dataset though, traditional functional connectivity analysis can be employed within subjects for future analyses. Functional connectivity analysis examines the correlation between brain activity in different regions within the same individual (Cole et al., 2014; Gonzalez-Castillo et al., 2015; Gonzalez-Castillo & Bandettini, 2018), allowing for the comparison of connectivity patterns across self, partner, and stranger conditions. This approach is similar to resting-state functional connectivity

methods but utilizes active stimuli, enabling the isolation of the self-condition and providing insights into the unique functional network architecture associated with processing one's own information. By comparing functional connectivity patterns across the three conditions, we can investigate how the functional network organization differs when processing self-related, close other-related, and stranger-related information. However, it is important to consider the potential influence of physiological noise on functional connectivity results, as autocorrelations within an individual's brain activity can introduce confounds (Birn et al., 2008; Van Dijk et al., 2010).

To complement the functional connectivity analysis and mitigate the issue of physiological noise, intersubject functional connectivity analysis can be employed. Intersubject functional connectivity quantifies the temporal correlation between brain activity in a seed region in one subject and activity in other brain regions in a different subject, effectively reducing the impact of physiological noise by examining correlations across subjects (Nastase et al., 2019). However, intersubject functional connectivity is still subject to the limitation of two-person neuroscience approaches in isolating the self-condition. Despite this limitation, comparing intersubject functional connectivity patterns between dyad and stranger conditions can provide valuable insights into the neural networks consistently engaged during the processing of self-relevant information. To overcome the limitations of both functional connectivity and intersubject functional connectivity approaches, combining the results of these analyses can be a powerful strategy, similar to the approach used with the univariate and ISC results in the current study. By identifying overlapping patterns of functional connectivity that are robust across both functional connectivity and intersubject functional connectivity methods, we can isolate the neural network architecture underlying self-referential processing that are consistent for both self and self-relevant information.

Dynamic functional connectivity analysis is a powerful technique that captures the temporal variations in functional connectivity patterns and allows for the investigation of moment-to-moment changes in network configurations (De Alteriis et al., 2024; Hutchison et al., 2013; Preti et al., 2017). In contrast to static functional connectivity, which assumes that the strength of connections between brain regions remains constant over time, dynamic functional connectivity analysis acknowledges the inherently dynamic nature of brain activity and connectivity (Calhoun et al., 2014). By employing sliding window approaches or time-frequency analyses, dynamic functional connectivity can reveal the temporal evolution of functional networks and provide insights into the dynamic interplay between different brain systems. In the context of self-referential processing, applying dynamic functional connectivity analysis to the narrative dataset would shed light on the temporal dynamics of the attention and salience networks and their interactions with sensory and default mode networks. This approach would help determine whether the observed differences in network connectivity across self, partner, and stranger conditions are sustained throughout the narrative or if there are momentary switches in network configurations. For example, dynamic functional connectivity analysis could reveal whether the heightened connectivity between attention and sensory networks during self-relevant narratives is a stable phenomenon or if it is characterized by transient episodes of increased coupling. Similarly, investigating the dynamic interplay between the salience network and default mode network could provide insights into the temporal coordination of these systems in response to self-relevant information. By capturing the moment-to-moment changes in functional connectivity, dynamic functional connectivity analysis would offer a more nuanced understanding of the neural dynamics underlying self-referential processing and help elucidate

the complex interplay between different brain networks in the context of the self as a powerful cognitive construct.

Combining dynamic functional connectivity analysis with the annotation of salient features in the narratives would be a logical next step then to further our understanding of the neural dynamics underlying self-referential processing. This method involves recruiting naive raters or employing computational methods to code the narratives for various aspects, such as emotional tone, surprising or important details, personal significance, and thematic content (Vodrahalli et al., 2018). Combining the annotated narratives with dynamic functional connectivity analysis would allow for a powerful investigation into the relationship between salient features of the stories and the observed switches in network architecture, providing insights into how different aspects of the narratives modulate the interplay between default mode, attention, and salience networks and how self-relevance influences the neural processing of salient information. By linking the dynamic changes in functional connectivity to specific, timestamped features of the narratives, we can gain a more mechanistic understanding of how the self, as a powerful cognitive construct, orchestrates the allocation of cognitive resources in response to personally relevant stimuli. The combination of functional connectivity, intersubject functional connectivity, dynamic functional connectivity, and narrative annotation, along with the previously discussed univariate and ISC findings, would provide a compelling body of evidence demonstrating the neural mechanisms through which the self influences cognitive processing.

Self-Relevance Drives Representational Modification

The findings from Chapters III and IV highlight the intricate interplay between generalized and individuated processing in the brain, particularly in the context of social cognition. Dual process learning theories provide a valuable framework for understanding how the brain navigates the balance between these two modes of processing (Ashby & Maddox, 2005; Love et al., 2004; Schapiro et al., 2017). According to these theories, the brain relies on both prototype-based (generalized) and exemplar-based (individuated) representations to efficiently process and store information (Love et al., 2004). Prototype-based representations involve the extraction of common features and the creation of a generalized, average representation of a category (Bowman et al., 2020; Reed, 1972). This mode of processing allows for rapid categorization and decision-making based on previously encountered information, making it highly efficient in situations where the stimuli share similar characteristics (Homa et al., 1981; Minda & Smith, 2001; Reed, 1972; Rosch & Mervis, 1975; J. D. Smith & Minda, 1998). In contrast, exemplar-based representations involve the storage of specific instances and their unique features, enabling the brain to make fine-grained distinctions and respond adaptively to novel situations (Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1986; Nosofsky & Johansen, 2000).

The brain's ability to utilize both generalized and individuated representations is crucial for navigating the complexities of the social world. In the context of person perception, the use of prototypical representations allows for rapid impression formation and the application of social categories, such as stereotypes (Brewer, 1988; S. T. Fiske & Neuberg, 1990). This mode of processing is particularly efficient when encountering strangers or individuals with whom we

have limited personal experience (Biesanz, 2021b; Chan et al., 2011b; Macrae & Bodenhausen, 2000). However, as social relationships develop and become more intimate, the brain likely shifts towards a more individuated processing style, focusing on the unique characteristics and experiences of the person (S. T. Fiske & Neuberg, 1990; Kunda & Thagard, 1996). This shift towards exemplar-based representations likely enables the formation of rich, multifaceted mental models of close others, facilitating empathy, perspective-taking, and the ability to anticipate their thoughts and behaviors (Decety & Jackson, 2004; Mitchell, Cloutier, et al., 2006). The findings from Chapters 3 and 4, which demonstrate the emergence of unique neural representations for close others and the influence of social closeness on the consistency of these representations, underscore the brain's flexibility in employing different processing strategies based on the social context and the depth of the relationship.

The concept of schemas, introduced by Piaget (1952), has been recognized as a fundamental organizing principle in cognitive psychology and has recently been linked to theories of prototype formation in neuroscience. The formation of prototypes in the brain involves the extraction of common features from a set of stimuli to create a generalized representation that captures the essential characteristics of a category (Reed, 1972; Rosch, 2002). The process of prototype adaptation, similar to Piaget's concepts of assimilation and accommodation, is still an active avenue of research. Assimilation occurs when new information is incorporated into existing schemas or prototypes without significantly altering their overall structure and accommodation involves the modification of schemas or prototypes to fit new information that cannot be easily integrated (Piaget, 1952). Recent work in neuroscience has highlighted the role of the hippocampus and the vmPFC in the formation and updating of prototypes (Bowman & Zeithamova, 2018; Schlichting & Preston, 2016; Zeithamova et al.,

2012), suggesting that these regions may be critical for the processes of assimilation and accommodation. The role of discriminative attention has emerged as a potential factor in driving these adaptive processes that modify existing generalized representational frameworks (Gottlieb, 2012; Musslick et al., 2020).

Discriminative attention, which involves the selective focusing of cognitive resources on the unique or distinguishing features of a stimulus (Chelazzi et al., 1993; Desimone & Duncan, 1995; Scholl, 2001; Treisman & Gelade, 1980), has been proposed as a key factor in facilitating the process of representational transformation (Musslick et al., 2020). When new information is encountered that is inconsistent with existing schemas, discriminative attention allows the brain to prioritize the specific features that challenge the current mental models (Beck & Kastner, 2009; Summerfield & Egner, 2009; van Kesteren et al., 2012, 2012). By directing cognitive resources towards these discrepancies, the brain can engage in a more detailed analysis of the information, updating and refining its schemas to better align with the new evidence. This attentional allocation likely serves as a catalyst for the accommodation process by enabling the brain to adapt its cognitive frameworks in response to novel or contradictory information. In the context of social cognition, discriminative attention likely plays a crucial role in shifting the brain's reliance away from generalized, prototype-based representations of others and towards more individuated, exemplar-based representations (S. T. Fiske & Neuberg, 1990; Kunda & Thagard, 1996). By focusing on the unique characteristics, behaviors, and experiences that define an individual, the brain can modify its existing schemas, creating rich, nuanced mental models that capture the complexity of close others and support the formation of deep, meaningful social connections (S. T. Fiske et al., 1999).

The heightened cognitive resources, particularly attention, that are recruited for processing self-relevant information may drive modulatory processes that modify existing representational structures, individuating them from the norm. This notion is supported by the findings from Chapter 3, which demonstrate that the vmPFC exhibited the strongest effect of social relationship strength on the uniqueness of neural representations for close others. The vmPFC's involvement is particularly interesting given its well-established role in self-referential processing (Denny et al., 2012; Macrae et al., 2004; Northoff et al., 2006; Rameson et al., 2010; Van Overwalle, 2009), valuation (Bartra et al., 2013; Rangel & Clithero, 2014), trait attribution (Kestemont et al., 2016; Ma, Baetens, Vandekerckhove, Van der Cruyssen, et al., 2014b; Van Overwalle, 2009), and mentalizing about close others (Gallagher & Frith, 2003; Mitchell, Neil Macrae, et al., 2005). Furthermore, the vmPFC has been implicated in the integration of features for inferential purposes (Bar, 2009; Euston et al., 2012; Kable & Glimcher, 2007; van Kesteren et al., 2012; Zeithamova et al., 2012), suggesting that it may play a crucial role in combining information in novel ways to create individuated representations. The unique representational patterns observed in the vmPFC for close others may reflect the outcome of a process in which the heightened cognitive resources recruited for self-relevant information are utilized to modify existing representations, driving them away from the norm. This finding highlights the potential mechanism through which self-relevance influences the neural representations of others, leading to the formation of individuated, person-specific concepts.

However, the behavioral results and the findings in the PCC suggest that the process of representational modification is not solely driven by self-relevance and discriminative attention. The endorsement of more normative traits for close others, despite the unique neural patterns in the vmPFC, indicates a possible influence of social norms and conventions on explicit

judgments. The PCC, with its involvement in autobiographical memory retrieval (Spreng et al., 2009; Svoboda et al., 2006) and the processing of socially relevant information (Leech & Sharp, 2013; Schilbach et al., 2008), may play a role in aligning an individual's responses with prevailing social expectations. In other words, while the vmPFC may be reflecting the true internal individuation of close others based on self-relevance and discriminative attention, the PCC may be involved in reconciling these unique representations with broader normative frameworks, guiding behavior to conform to social norms.

The findings from Chapter IV demonstrate that the default mode network as a whole contains reliable social information that can differentiate group members from each other above chance and irrespective of the endorsement of the trait on which the individual was being evaluated. This result represents a significant advancement in person decoding research, as previous studies have primarily focused on decoding stimulus-driven features such as face or voice (Anzellotti & Caramazza, 2017; Kriegeskorte et al., 2007). The successful decoding of person identity based on abstract trait representations within the default mode network highlights the richness and specificity of the social information processed in these regions, extending our understanding of the neural basis of person perception and social cognition. However, the low accuracy of the model raises questions about the consistency and stability of the unique representations observed in Chapter III. It is possible that the individuation process, driven by self-relevance and discriminative attention, may result in trial-specific and variable representations that are challenging to decode consistently.

Interestingly, the sensitivity analysis in Chapter IV revealed an inverse relationship between social closeness and decoding accuracy in the PCC, suggesting that, at least in the PCC,

it may be easier to decode representations of others who are not close to the self. One possible explanation is that non-close others are still represented using generalized, normative representations that remain stable across trials, while the individualized representations of close others, shaped by self-relevance and discriminative attention, may be more dynamic and context-dependent, varying as different aspects of an individual's identity become salient. Future research should aim to disentangle these regional differences and investigate the factors that contribute to the stability and reliability of social representations in the default mode network.

While the findings from Chapters III and IV provide evidence for the role of self-relevance in shaping the neural representations of others, it is important to acknowledge the limitations in interpreting these results. The interpretation of these findings as evidence for the involvement of attentional processes and representational modification is based primarily on theoretical arguments, previous literature, and the logical progression from the results of Chapter II, which demonstrated the recruitment of cognitive processes such as attention by self-relevant material. However, it is crucial to note that the study used in Chapters III and IV was not specifically designed to investigate attention or memory processes. The trait adjective task employed in the study, where participants simply responded “yes” or “no” to whether a trait accurately described a target, does not directly measure or manipulate attention or examine learning and memory processes. Consequently, the links between self-relevance, discriminative attention, and representational modification, as well as the mechanisms of assimilation and accommodation, are inferred rather than directly tested. This limitation in the experimental design constrains our ability to draw definitive conclusions about the specific cognitive processes underlying the observed neural representations.

To address these limitations, future research should focus on designing experiments that directly manipulate and measure attentional processes. This could involve employing eye-tracking or other attentional measures to assess the allocation of discriminative attention to specific features of social stimuli or manipulating the salience or relevance of social information to examine the impact on neural representations. Investigating the role of learning and memory processes in shaping neural representations is also crucial. This could be achieved through longitudinal studies that track the formation and modification of neural representations over time, or by employing experimental paradigms that directly test the mechanisms of assimilation and accommodation, such as schema-consistent and schema-inconsistent information processing. By addressing these limitations and pursuing these future directions, researchers can strengthen the evidence for the role of self-relevance in driving discriminative attention and representational modification, ultimately contributing to a more comprehensive understanding of the neural basis of person perception and social cognition.

Conclusion

In conclusion, this dissertation provides evidence for the profound impact of self-relevant information on neural processing and cognitive resource allocation. The overarching findings demonstrate that self-related information has the power to dramatically alter the way the brain deploys cognitive resources, particularly in social contexts. These studies together imply that self-relevant information results in the recruitment of attention mechanisms that can enhance the signal-to-noise ratio of incoming sensory information, promoting deeper processing. This attentional allocation can serve as a catalyst for representational modification, whereby important information is scrutinized for details that may be consistent or inconsistent with existing

schematic structures. In regard to our perceptions of close others, this process can lead to the individuation of normative structures, resulting in the creation of person-specific schemas that are more context-dependent and dynamic.

By employing a suite of contemporary neuroimaging methods, multivariate computational approaches, and using novel experimental paradigms, this work attempts to move theoretical insights from philosophy and cognitive science towards a more mechanistic explanation of self-referential processing. This research contributes to a more nuanced and integrative theory of the self, emphasizing its central role in shaping our perceptions, thoughts, and behaviors in the social world. It underscores the functional role of the self as a powerful cognitive construct that not only influences the processing of personally significant information but also modulates our understanding and representation of others. While further research is needed to directly examine the specific mechanisms of attention, learning, and memory and the network dynamics that are involved, this work lays a foundation for future investigations into the nature of self-representation and its impact on social cognition. Ultimately, this dissertation advances our understanding of how the self, as a central organizing principle in human cognition, guides the construction of our conscious experience and shapes our interactions with the social world.

References

Adolphs, R. (2009). The social brain: Neural basis of social knowledge. *Annual Review of Psychology, 60*, 693–716. https://doi.org/10.1146/annurev.psych.60.110707.163514

Alexopoulos, T., Muller, D., Ric, F., & Marendaz, C. (2012). I, me, mine: Automatic attentional capture by self-related stimuli. *European Journal of Social Psychology, 42*(6), 770–779. https://doi.org/10.1002/ejsp.1882

Alicke, M. D., & Sedikides, C. (2009). Self-enhancement and self-protection: What they are and what they do. *European Review of Social Psychology, 20*, 1–48. https://doi.org/10.1080/10463280802613866

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

Alves, P. N., Foulon, C., Karolis, V., Bzdok, D., Margulies, D. S., Volle, E., & Thiebaut de Schotten, M. (2019). An improved neuroanatomical model of the default-mode network reconciles previous neuroimaging and neuropathological findings. *Communications Biology, 2*(1), Article 1. https://doi.org/10.1038/s42003-019-0611-3

Amaro, E., & Barker, G. J. (2006). Study design in fMRI: Basic principles. *Brain and Cognition, 60*(3), 220–232. https://doi.org/10.1016/j.bandc.2005.11.009

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

Anderson, N. H. (1968). Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology, 9*(3), 272–279. https://doi.org/10.1037/h0025907

Andrews-Hanna, J. R., Kaiser, R. H., Turner, A. E. J., Reineberg, A. E., Godinez, D., Dimidjian, S., & Banich, M. T. (2013). A penny for your thoughts: Dimensions of self-generated thought content and relationships with individual differences in emotional wellbeing. *Frontiers in Psychology*, *4*, 900. https://doi.org/10.3389/fpsyg.2013.00900

Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., & Buckner, R. L. (2010). Functional-Anatomic Fractionation of the Brain's Default Network. *Neuron*, *65*(4), 550–562. https://doi.org/10.1016/j.neuron.2010.02.005

Andrews-Hanna, J. R., Smallwood, J., & Spreng, R. N. (2014). The default network and self-generated thought: Component processes, dynamic control, and clinical relevance. *Annals of the New York Academy of Sciences*, *1316*, 29–52. https://doi.org/10.1111/nyas.12360

Anzellotti, S., & Caramazza, A. (2017). Multimodal representations of person identity individuated with fMRI. *Cortex*, *89*, 85–97. https://doi.org/10.1016/j.cortex.2017.01.013

Anzellotti, S., Fairhall, S. L., & Caramazza, A. (2014). Decoding Representations of Face Identity That are Tolerant to Rotation. *Cerebral Cortex*, *24*(8), 1988–1995. https://doi.org/10.1093/cercor/bht046

Arioli, M., Gianelli, C., & Canessa, N. (2021). Neural representation of social concepts: A coordinate-based meta-analysis of fMRI studies. *Brain Imaging and Behavior*, *15*(4), 1912–1921. https://doi.org/10.1007/s11682-020-00384-6

Aristotle. (1999). *Nicomachean Ethics* (M. Ostwald, Trans.; 1st edition). Pearson.

Aron, A., Aron, E. N., Tudor, M., & Nelson, G. (1991). Close relationships as including other in the self. *Journal of Personality and Social Psychology*, *60*(2), 241–253. https://doi.org/10.1037/0022-3514.60.2.241

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology, 56*, 149–178. https://doi.org/10.1146/annurev.psych.56.091103.070217

Audio recording and editing software | Adobe Audition. (n.d.). Retrieved June 18, 2024, from https://www.adobe.com/products/audition.html

Axelrod, V., & Yovel, G. (2015). Successful Decoding of Famous Faces in the Fusiform Face Area. *PLOS ONE*, *10*(2), e0117126. https://doi.org/10.1371/journal.pone.0117126

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron*, *95*(3), 709-721.e5. https://doi.org/10.1016/j.neuron.2017.06.041

Baldassano, C., Hasson, U., & Norman, K. A. (2018). Representation of Real-World Event Schemas during Narrative Perception. *The Journal of Neuroscience*, *38*(45), 9689–9699. https://doi.org/10.1523/JNEUROSCI.0251-18.2018

Bar, M. (2009). The proactive brain: Memory for predictions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1235–1243. https://doi.org/10.1098/rstb.2008.0310

Barnett, B. O., Brooks, J. A., & Freeman, J. B. (2021). Stereotypes bias face perception via orbitofrontal–fusiform cortical interaction. *Social Cognitive and Affective Neuroscience*, *16*(3), 302–314. https://doi.org/10.1093/scan/nsaa165

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage, 76*, 412–427. https://doi.org/10.1016/j.neuroimage.2013.02.063

Beauvoir, S. de, Parshley, H. M., & Beauvoir, S. de. (1989). *The second sex*. Vintage Books.

Beck, D. M., & Kastner, S. (2009). Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vision Research*, *49*(10), 1154–1165. https://doi.org/10.1016/j.visres.2008.07.012

Beer, J. S., Heerey, E. A., Keltner, D., Scabini, D., & Knight, R. T. (2003). The regulatory function of self-conscious emotion: Insights from patients with orbitofrontal damage. *Journal of Personality and Social Psychology*, *85*(4), 594–604. https://doi.org/10.1037/0022-3514.85.4.594

Beer, J. S., & Hughes, B. L. (2010). Neural systems of social comparison and the "above-average" effect. *NeuroImage*, *49*(3), 2671–2679. https://doi.org/10.1016/j.neuroimage.2009.10.075

Beer, J. S., Lombardo, M. V., & Bhanji, J. P. (2010). Roles of medial prefrontal cortex and orbitofrontal cortex in self-evaluation. *Journal of Cognitive Neuroscience*, *22*(9), 2108–2119. https://doi.org/10.1162/jocn.2009.21359

Bendor, D., & Wang, X. (2006). Cortical representations of pitch in monkeys and humans. *Current Opinion in Neurobiology*, *16*(4), 391–399. https://doi.org/10.1016/j.conb.2006.07.001

Ben-Yakov, A., Honey, C. J., Lerner, Y., & Hasson, U. (2012). Loss of reliable temporal structure in event-related averaging of naturalistic stimuli. *NeuroImage*, *63*(1), 501–506. https://doi.org/10.1016/j.neuroimage.2012.07.008

Berkman, E. T., Livingston, J. L., & Kahn, L. E. (2017). Finding the "self" in self-regulation: The identity-value model. *Psychological Inquiry*, *28*(2–3), 77–98. https://doi.org/10.1080/1047840X.2017.1323463

Biesanz, J. C. (2021a). The Social Accuracy Model. In T. D. Letzring & J. S. Spain (Eds.), *The Oxford Handbook of Accurate Personality Judgment* (pp. 60–82). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190912529.013.5

Biesanz, J. C. (2021b). The Social Accuracy Model. In T. D. Letzring & J. S. Spain (Eds.), *The Oxford Handbook of Accurate Personality Judgment* (pp. 60–82). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190912529.013.5

Biesanz, J. C., West, S. G., & Millevoi, A. (2007). What do you learn about someone over time? The relationship between length of acquaintance and consensus and self-other agreement in judgments of personality. *Journal of Personality and Social Psychology*, *92*(1), 119–135. https://doi.org/10.1037/0022-3514.92.1.119

Birn, R. M., Smith, M. A., Jones, T. B., & Bandettini, P. A. (2008). The Respiration Response Function: The temporal dynamics of fMRI signal fluctuations related to changes in respiration. *NeuroImage*, *40*(2), 644–654. https://doi.org/10.1016/j.neuroimage.2007.11.059

Blackford, J. U., Avery, S. N., Cowan, R. L., Shelton, R. C., & Zald, D. H. (2011). Sustained amygdala response to both novel and newly familiar faces characterizes inhibited temperament. *Social Cognitive and Affective Neuroscience*, *6*(5), 621–629. https://doi.org/10.1093/scan/nsq073

Bluck, S., & Alea, N. (2009). Thinking and talking about the past: Why remember? *Applied Cognitive Psychology*, *23*(8), 1089–1104. https://doi.org/10.1002/acp.1612

Bower, G. H., & Gilligan, S. G. (1979). Remembering information related to one's self. *Journal of Research in Personality*, *13*(4), 420–432. https://doi.org/10.1016/0092-6566(79)90005-9

Bowman, C. R., Iwashita, T., & Zeithamova, D. (2020). Tracking prototype and exemplar representations in the brain across learning. *eLife*, *9*, e59360. https://doi.org/10.7554/eLife.59360

Bowman, C. R., & Zeithamova, D. (2018). Abstract Memory Representations in the Ventromedial Prefrontal Cortex and Hippocampus Support Concept Generalization. *The Journal of Neuroscience*, *38*(10), 2605–2614. https://doi.org/10.1523/JNEUROSCI.2811-17.2018

Brefczynski, J. A., & DeYoe, E. A. (1999). A physiological correlate of the "spotlight" of visual attention. *Nature Neuroscience*, *2*(4), 370–374. https://doi.org/10.1038/7280

Brewer, M. B. (1988). A dual process model of impression formation. In *A dual process model of impression formation* (pp. 1–36). Lawrence Erlbaum Associates, Inc.

Brooks, J. A., & Freeman, J. B. (2019). Neuroimaging of person perception: A social-visual interface. *Neuroscience Letters*, *693*, 40–43. https://doi.org/10.1016/j.neulet.2017.12.046

Brown, C. L., Chen, K.-H., Wells, J. L., Otero, M. C., Connelly, D. E., Levenson, R. W., & Fredrickson, B. L. (2022). Shared emotions in shared lives: Moments of co-experienced affect, more than individually experienced affect, linked to relationship quality. *Emotion*, *22*(6), 1387–1393. https://doi.org/10.1037/emo0000939

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

Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences*, *11*(2), 49–57. https://doi.org/10.1016/j.tics.2006.11.004

Buckner, R. L., & DiNicola, L. M. (2019). The brain's default network: Updated anatomy, physiology and evolving insights. *Nature Reviews. Neuroscience*, *20*(10), 593–608. https://doi.org/10.1038/s41583-019-0212-7

Butler, J. (2006). *Gender trouble: Feminism and the subversion of identity*. Routledge.

Calhoun, V. D., Miller, R., Pearlson, G., & Adalı, T. (2014). The Chronnectome: Time-Varying Connectivity Networks as the Next Frontier in fMRI Data Discovery. *Neuron*, *84*(2), 262–274. https://doi.org/10.1016/j.neuron.2014.10.015

Carlin, J. D. (2015). Decoding Face Exemplars from fMRI Responses: What Works, What Doesn't? *Journal of Neuroscience*, *35*(25), 9252–9254. https://doi.org/10.1523/JNEUROSCI.1385-15.2015

Carter, R. M., & Huettel, S. A. (2013). A nexus model of the temporal-parietal junction. In *Trends in Cognitive Sciences* (Vol. 17). https://doi.org/10.1016/j.tics.2013.05.007

Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: A review of its functional anatomy and behavioural correlates. *Brain*, *129*(3), 564–583.

Chan, M., Rogers, K. H., Parisotto, K. L., & Biesanz, J. C. (2011a). Forming first impressions: The role of gender and normative accuracy in personality perception. *Journal of Research in Personality*, *45*(1), 117–120. https://doi.org/10.1016/j.jrp.2010.11.001

Chan, M., Rogers, K. H., Parisotto, K. L., & Biesanz, J. C. (2011b). Forming first impressions: The role of gender and normative accuracy in personality perception. *Journal of Research in Personality*, *45*(1), 117–120. https://doi.org/10.1016/j.jrp.2010.11.001

Chang, L. J., Yarkoni, T., Khaw, M. W., & Sanfey, A. G. (2013). Decoding the role of the insula in human cognition: Functional parcellation and large-scale reverse inference. *Cerebral Cortex$\backslash$*, *23*(3), 739–749. https://doi.org/10.1093/cercor/bhs065

Chaumon, M., Kveraga, K., Barrett, L. F., & Bar, M. (2014). Visual Predictions in the Orbitofrontal Cortex Rely on Associative Content. *Cerebral Cortex*, *24*(11), 2899–2907. https://doi.org/10.1093/cercor/bht146

Chavez, R. S. (2021). Tangled Representations of Self and Others in the Medial Prefrontal Cortex. In M. Gilead & K. N. Ochsner (Eds.), *The Neural Basis of Mentalizing* (pp. 599–611). Springer International Publishing. https://doi.org/10.1007/978-3-030-51890-5_31

Chavez, R. S., & Heatherton, T. F. (2015). Multimodal frontostriatal connectivity underlies individual differences in self-esteem. *Social Cognitive and Affective Neuroscience*, *10*(3), 364–370. https://doi.org/10.1093/scan/nsu063

Chavez, R. S., Heatherton, T. F., & Wagner, D. D. (2017). Neural Population Decoding Reveals the Intrinsic Positivity of the Self. *Cerebral Cortex*, *27*(11), 5222–5229. https://doi.org/10.1093/cercor/bhw302

Chavez, R. S., & Wagner, D. D. (2020). The neural representation of self is recapitulated in the brains of friends: A round-robin fMRI study. *Journal of Personality and Social Psychology*, *118*(3), 407–416. https://doi.org/10.1037/pspa0000178

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature*, *363*(6427), 345–347. https://doi.org/10.1038/363345a0

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017a). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, *20*(1), 115–125. https://doi.org/10.1038/nn.4450

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017b). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, *20*(1). https://doi.org/10.1038/nn.4450

Cheng, X., Popal, H., Wang, H., Hu, R., Zang, Y., Zhang, M., Thornton, M. A., Cai, H., Bi, Y., Reilly, J., Olson, I. R., & Wang, Y. (2023). *The Conceptual Structure of Human Relationships Across Modern and Historical Cultures*. https://doi.org/10.31234/osf.io/ut6qp

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, *25*, 975–979. https://doi.org/10.1121/1.1907229

Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., & Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(21), 8719–8724. https://doi.org/10.1073/pnas.0900234106

Cikara, M., Van Bavel, J. J., Ingbretsen, Z. A., & Lau, T. (2017). Decoding "us" and "them": Neural representations of generalized group concepts. *Journal of Experimental Psychology: General*, *146*(5), 621–631. https://doi.org/10.1037/xge0000287

Cisek, P. (2019). Resynthesizing behavior through phylogenetic refinement. *Attention, Perception, & Psychophysics*, *81*(7), 2265–2287. https://doi.org/10.3758/s13414-019-01760-1

Cloutier, J., Gabrieli, J. D. E., O'Young, D., & Ambady, N. (2011). An fMRI study of violations of social expectations: When people are not who we expect them to be. *NeuroImage*, *57*(2), 583–588. https://doi.org/10.1016/j.neuroimage.2011.04.051

Cole, M. W., Bassett, D. S., Power, J. D., Braver, T. S., & Petersen, S. E. (2014). Intrinsic and task-evoked network architectures of the human brain. *Neuron*, *83*(1), 238–251. https://doi.org/10.1016/j.neuron.2014.05.014

Contreras, J. M., Banaji, M. R., & Mitchell, J. P. (2013). Multivoxel Patterns in Fusiform Face Area Differentiate Faces by Sex and Race. *PLoS ONE*, *8*(7), e69684. https://doi.org/10.1371/journal.pone.0069684

Conway, M. A., Pothos, E. M., & Turk, D. J. (2016). The self-relevance system? *Cognitive Neuroscience*, *7*(1–4), 20–21. https://doi.org/10.1080/17588928.2015.1075484

Cosme, D., Flournoy, J., Livingston, J., Lieberman, M., Dapretto, M., & Pfeifer, J. (2021). *Testing the adolescent social reorientation model during self and other evaluation using hierarchical growth curve modeling with parcellated fMRI data*. PsyArXiv. https://doi.org/10.31234/osf.io/8eyf5

Courtney, A. L., & Meyer, M. L. (2020). Self-Other Representation in the Social Brain Reflects Social Connection. *The Journal of Neuroscience*, *40*(29), 5616–5627. https://doi.org/10.1523/JNEUROSCI.2826-19.2020

Craig, A. D. B. (2011). Significance of the insula for the evolution of human awareness of feelings from the body. *Annals of the New York Academy of Sciences*, *1225*, 72–82. https://doi.org/10.1111/j.1749-6632.2011.05990.x

Craik, F. I. M., Moroz, T. M., Moscovitch, M., Stuss, D. T., Winocur, G., Tulving, E., & Kapur, S. (1999). In Search of the Self: A Positron Emission Tomography Study. *Psychological Science*, *10*(1), 26–34. https://doi.org/10.1111/1467-9280.00102

Cronbach, L. J. (1955). Processes affecting scores on “understanding of others” and “assumed similarity.” *Psychological Bulletin*, *52*(3), 177–193. https://doi.org/10.1037/h0044919

Crone, E. A., & Dahl, R. E. (2012). Understanding adolescence as a period of social–affective engagement and goal flexibility. *Nature Reviews Neuroscience*, *13*(9), 636–650. https://doi.org/10.1038/nrn3313

Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping*, *8*(2–3), 109–114. https://doi.org/10.1002/(SICI)1097-0193(1999)8:2/3<109::AID-HBM7>3.0.CO;2-W

Damasio, A. R., Tranel, D., & Damasio, H. (1990). Individuals with sociopathic behavior caused by frontal damage fail to respond autonomically to social stimuli. *Behavioural Brain Research*, *41*(2), 81–94. https://doi.org/10.1016/0166-4328(90)90144-4

Dang, T. P., Mattan, B. D., Kubota, J. T., & Cloutier, J. (2019). The ventromedial prefrontal cortex is particularly responsive to social evaluations requiring the use of person-knowledge. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-41544-z

D'Argembeau, A., Cassol, H., Phillips, C., Balteau, E., Salmon, E., & Van der Linden, M. (2014). Brains creating stories of selves: The neural basis of autobiographical reasoning. *Social Cognitive and Affective Neuroscience*, *9*(5), 646–652. https://doi.org/10.1093/scan/nst028

D'Argembeau, A., Jedidi, H., Balteau, E., Bahri, M., Phillips, C., & Salmon, E. (2012). Valuing one's self: Medial prefrontal involvement in epistemic and emotive investments in self-views. *Cerebral Cortex*, *22*(3), 659–667.

D'Argembeau, A., Renaud, O., & Van Der Linden, M. (2011). Frequency, characteristics and functions of future-oriented thoughts in daily life. *Applied Cognitive Psychology*, *25*(1), 96–103. https://doi.org/10.1002/acp.1647

Davis, M., & Whalen, P. J. (2001). The amygdala: Vigilance and emotion. *Molecular Psychiatry*, *6*(1), 13–34.

De Alteriis, G., Sherwood, O., Ciaramella, A., Leech, R., Cabral, J., Turkheimer, F. E., & Expert, P. (2024). *DySCo: A general framework for dynamic Functional Connectivity*. https://doi.org/10.1101/2024.06.12.598743

De Felice, S., Hakim, U., Gunasekara, N., Pinti, P., Tachtsidis, I., & Hamilton, A. (2024). Having a chat and then watching a movie: How social interaction synchronises our brains during co-watching. *Oxford Open Neuroscience*, *3*, kvae006. https://doi.org/10.1093/oons/kvae006

Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, *3*(2), 71–100. https://doi.org/10.1177/1534582304267187

Denny, B. T., Kober, H., Wager, T. D., & Ochsner, K. N. (2012). A meta-analysis of functional neuroimaging studies of self- and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *Journal of Cognitive Neuroscience*, *24*(8), 1742–1752. https://doi.org/10.1162/jocn_a_00233

Descartes, R., & Cottingham, J. (2013). *Meditations on first philosophy: With selections from the objections and replies ; a Latin-English edition*. Cambridge University Pr.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205

DiQuattro, N. E., & Geng, J. J. (2011). Contextual knowledge configures attentional control networks. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(49), 18026–18035. https://doi.org/10.1523/JNEUROSCI.4040-11.2011

Dumontheil, I., Küster, O., Apperly, I. A., & Blakemore, S.-J. (2010). Taking perspective into account in a communicative task. *NeuroImage*, *52*(4), 1574–1583. https://doi.org/10.1016/j.neuroimage.2010.05.056

Eichenbaum, H. (2004). Hippocampus: Cognitive processes and neural representations that underlie declarative memory. *Neuron*, *44*(1), 109–120. https://doi.org/10.1016/j.neuron.2004.08.028

Eippert, F., Veit, R., Weiskopf, N., Erb, M., Birbaumer, N., & Anders, S. (2007). Regulation of emotional responses elicited by threat-related stimuli. *Human Brain Mapping*, *28*(5), 409–423. https://doi.org/10.1002/hbm.20291

Erikson, E. H. (1994). *Identity and the life cycle* (Reissued as Norton paperback). W. W. Norton & Company.

Euston, D. R., Gruber, A. J., & McNaughton, B. L. (2012). The role of medial prefrontal cortex in memory and decision making. *Neuron*, *76*(6), 1057–1070.

Fareri, D. S., Niznikiewicz, M. A., Lee, V. K., & Delgado, M. R. (2012). Social Network Modulation of Reward-Related Signals. *Journal of Neuroscience*, *32*(26), 9045–9052. https://doi.org/10.1523/JNEUROSCI.0610-12.2012

Fiske, A. P., & Haslam, N. (1996). Social Cognition Is Thinking About Relationships. *Current Directions in Psychological Science*, *5*(5), 143–148. https://doi.org/10.1111/1467-8721.ep11512349

Fiske, S. T., Lin, M., & Neuberg, S. L. (1999). The continuum model: Ten years later. In *Dual-process theories in social psychology* (In S. Chaiken&Y. Trope (Eds.), pp. 231–254). Guilford Press.

Fiske, S. T., & Neuberg, S. L. (1990). A Continuum of Impression Formation, from Category-Based to Individuating Processes: Influences of Information and Motivation on Attention and Interpretation. In *Advances in Experimental Social Psychology* (Vol. 23, pp. 1–74). Elsevier. https://doi.org/10.1016/S0065-2601(08)60317-2

Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). "Who" is saying "what"? Brain-based decoding of human voice and speech. *Science (New York, N.Y.)*, *322*(5903), 970–973. https://doi.org/10.1126/science.1164318

Freeman, J. B., Johnson, K. L., Adams, R. B., & Ambady, N. (2012). The social-sensory interface: Category interactions in person perception. *Frontiers in Integrative Neuroscience*, *6*. https://doi.org/10.3389/fnint.2012.00081

Freeman, J. B., Schiller, D., Rule, N. O., & Ambady, N. (2010). The neural origins of superficial and individuated judgments about ingroup and outgroup members. *Human Brain Mapping*, *31*(1), 150–159. https://doi.org/10.1002/hbm.20852

Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Auditory attention—Focusing the searchlight on sound. *Current Opinion in Neurobiology*, *17*(4), 437–455. https://doi.org/10.1016/j.conb.2007.07.011

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

Golland, Y., Arzouan, Y., & Levit-Binnun, N. (2015). The Mere Co-Presence: Synchronization of Autonomic Signals and Emotional Responses across Co-Present Individuals Not

Engaged in Direct Interaction. *PLOS ONE, 10*(5), e0125804. https://doi.org/10.1371/journal.pone.0125804

Gonzalez-Castillo, J., & Bandettini, P. A. (2018). Task-based Dynamic Functional Connectivity: Recent findings and open questions. *NeuroImage, 180*(Pt B), 526–533. https://doi.org/10.1016/j.neuroimage.2017.08.006

Gonzalez-Castillo, J., Hoy, C. W., Handwerker, D. A., Robinson, M. E., Buchanan, L. C., Saad, Z. S., & Bandettini, P. A. (2015). Tracking ongoing cognition in individuals using brief, whole-brain functional connectivity patterns. *Proceedings of the National Academy of Sciences*, *112*(28), 8762–8767. https://doi.org/10.1073/pnas.1501242112

Gottlieb, J. (2012). Attention, Learning, and the Value of Information. *Neuron*, *76*(2), 281–295. https://doi.org/10.1016/j.neuron.2012.09.034

Grady, C. L., Van Meter, J. W., Maisog, J. M., Pietrini, P., Krasuski, J., & Rauschecker, J. P. (1997). Attention-related modulation of activity in primary and secondary auditory cortex. *Neuroreport*, *8*(11), 2511–2516. https://doi.org/10.1097/00001756-199707280-00019

Greenwald, A. G., & Banaji, M. R. (1989). The self as a memory system: Powerful, but ordinary. In *Journal of Personality and Social Psychology* (Vol. 57).

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, *15*(8), 536–548. https://doi.org/10.1038/nrn3747

Guthrie, T. D., Benadjaoud, Y. Y., & Chavez, R. S. (2022). Social Relationship Strength Modulates the Similarity of Brain-to-Brain Representations of Group Members. *Cerebral Cortex*, *32*(11), 2469–2477. https://doi.org/10.1093/cercor/bhab355

Hare, T. A., Camerer, C. F., Knoepfle, D. T., O'Doherty, J. P., & Rangel, A. (2010). Value Computations in Ventral Medial Prefrontal Cortex during Charitable Decision Making Incorporate Input from Regions Involved in Social Cognition. *Journal of Neuroscience*, *30*(2), 583–590. https://doi.org/10.1523/JNEUROSCI.4089-09.2010

Harris, L. T., McClure, S. M., Van Den Bos, W., Cohen, J. D., & Fiske, S. T. (2007). Regions of the MPFC differentially tuned to social and nonsocial affective evaluation. *Cognitive, Affective, & Behavioral Neuroscience*, *7*(4), 309–316. https://doi.org/10.3758/CABN.7.4.309

Harris, L. T., Todorov, A., & Fiske, S. T. (2005). Attributions on the brain: Neuro-imaging dispositional inferences, beyond theory of mind. *NeuroImage*, *28*(4), 763–769. https://doi.org/10.1016/j.neuroimage.2005.05.021

Haslam, N. (2006). Dehumanization: An Integrative Review. *Personality and Social Psychology Review*, *10*(3), 252–264. https://doi.org/10.1207/s15327957pspr1003_4

Hassabis, D., Spreng, R. N., Rusu, A. a, Robbins, C. a, Mar, R. a, & Schacter, D. L. (2014). Imagine All the People: How the Brain Creates and Uses Personality Models to Predict Behavior. *Cerebral Cortex*, *24*(8), 1979–1987. https://doi.org/10.1093/cercor/bht042

Hasson, U. (2004). Intersubject Synchronization of Cortical Activity During Natural Vision. *Science*, *303*(5664), 1634–1640. https://doi.org/10.1126/science.1089506

Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: Memory as an integral component of information processing. *Trends in Cognitive Sciences*, *19*(6), 304–313. https://doi.org/10.1016/j.tics.2015.04.006

Hasson, U., Malach, R., & Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends in Cognitive Sciences*, *14*(1), 40–48. https://doi.org/10.1016/j.tics.2009.10.011

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A Hierarchy of Temporal Receptive Windows in Human Cortex. *Journal of Neuroscience*, *28*(10), 2539–2550. https://doi.org/10.1523/JNEUROSCI.5487-07.2008

Haxby, J. V. (2012). Multivariate pattern analysis of fMRI: the early beginnings. *NeuroImage*, *62*(2), 852–855.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science*, *293*(5539), 2425–2430. https://doi.org/10.1126/science.1063736

Heatherton, T. F. (2011). Neuroscience of self and self-regulation. *Annual Review of Psychology*, *62*, 363–390.

Heatherton, T. F., Wyland, C. L., Macrae, C. N., Demos, K. E., Denny, B. T., & Kelley, W. M. (2006). Medial prefrontal activity differentiates self from close others. *Social Cognitive and Affective Neuroscience*, *1*(1), 18–25. https://doi.org/10.1093/scan/nsl001

Heleven, E., & Van Overwalle, F. (2019). Neural representations of others in the medial prefrontal cortex do not depend on our knowledge about them. *Social Neuroscience*, *14*(3), 286–299. https://doi.org/10.1080/17470919.2018.1472139

Holland, A. C., & Kensinger, E. A. (2010). Emotion and autobiographical memory. *Physics of Life Reviews, 7*(1), 88–131. https://doi.org/10.1016/j.plrev.2010.01.006

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory, 7*(6), 418–439. https://doi.org/10.1037/0278-7393.7.6.418

Honey, C. J., Thesen, T., Donner, T. H., Silbert, L. J., Carlson, C. E., Devinsky, O., Doyle, W. K., Rubin, N., Heeger, D. J., & Hasson, U. (2012). Slow Cortical Dynamics and the Accumulation of Information over Long Timescales. *Neuron, 76*(2), 423–434. https://doi.org/10.1016/j.neuron.2012.08.011

Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nature Neuroscience, 3*(3), 284–291. https://doi.org/10.1038/72999

Hornak, J., Bramham, J., Rolls, E. T., Morris, R. G., O'Doherty, J., Bullock, P. R., & Polkey, C. E. (2003). Changes in emotion after circumscribed surgical lesions of the orbitofrontal and cingulate cortices. *Brain: A Journal of Neurology, 126*(Pt 7), 1691–1712. https://doi.org/10.1093/brain/awg168

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology, 195*(1), 215–243. https://doi.org/10.1113/jphysiol.1968.sp008455

Hughes, B. L., & Beer, J. S. (2013). Protecting the self: The effect of social-evaluative threat on neural representations of self. *Journal of Cognitive Neuroscience, 25*(4), 613–622.

Human, L. J., Carlson, E. N., Geukes, K., Nestler, S., & Back, M. D. (2020). Do accurate personality impressions benefit early relationship development? The bidirectional

associations between accuracy and liking. *Journal of Personality and Social Psychology*, *118*(1), 199–212. https://doi.org/10.1037/pspp0000214

Humphreys, G. W., & Sui, J. (2016). Attentional control and the self: The Self-Attention Network (SAN). *Cognitive Neuroscience*, *7*(1–4), 5–17. https://doi.org/10.1080/17588928.2015.1044427

Humphries, C., Liebenthal, E., & Binder, J. R. (2010). Tonotopic organization of human auditory cortex. *NeuroImage*, *50*(3), 1202–1211. https://doi.org/10.1016/j.neuroimage.2010.01.046

Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., Penna, S. D., Duyn, J. H., Glover, G. H., Gonzalez-Castillo, J., Handwerker, D. A., Keilholz, S., Kiviniemi, V., Leopold, D. A., de Pasquale, F., Sporns, O., Walter, M., & Chang, C. (2013). Dynamic functional connectivity: Promise, issues, and interpretations. *NeuroImage*, *80*, 10.1016/j.neuroimage.2013.05.079. https://doi.org/10.1016/j.neuroimage.2013.05.079

Jääskeläinen, I. P., & Kosonogov, V. (2023). Perspective taking in the human brain: Complementary evidence from neuroimaging studies with media-based naturalistic stimuli and artificial controlled paradigms. *Frontiers in Human Neuroscience*, *17*, 1051934. https://doi.org/10.3389/fnhum.2023.1051934

James, W. (1890). *The Principles of Psychology* (Vol. 1). Holt.

Jäncke, L., Mirzazade, S., & Joni Shah, N. (1999). Attention modulates activity in the primary and the secondary auditory cortex: A functional magnetic resonance imaging study in human subjects. *Neuroscience Letters*, *266*(2), 125–128. https://doi.org/10.1016/S0304-3940(99)00288-8

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, *62*(2), 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience*, *10*(12), 1625–1633.

Kahneman, D., Treisman, A., & Burkell, J. (1983). The cost of visual filtering. *Journal of Experimental Psychology: Human Perception and Performance*, *9*(4), 510–522. https://doi.org/10.1037/0096-1523.9.4.510

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, *8*(5), 679–685. https://doi.org/10.1038/nn1444

Kant, I., Guyer, P., & Wood, A. W. (2000). *Critique of pure reason*. Cambridge university press.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *361*(1476), 2109–2128. https://doi.org/10.1098/rstb.2006.1934

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron*, *22*(4), 751–761. https://doi.org/10.1016/s0896-6273(00)80734-5

Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002). Finding the self? An event-related fMRI study. *Journal of Cognitive Neuroscience*, *14*(5), 785–794.

Kestemont, J., Van Mieghem, A., Beeckmans, K., Van Overwalle, F., & Vandekerckhove, M. (2016). Social attributions in patients with ventromedial prefrontal hypoperfusion. *Social Cognitive and Affective Neuroscience*, *11*(4), 652–662. https://doi.org/10.1093/scan/nsv147

Keyes, H., & Brady, N. (2010). Self-face recognition is characterized by “bilateral gain” and by faster, more accurate performance which persists when faces are inverted. *The Quarterly Journal of Experimental Psychology*, *63*(5), 840–847. https://doi.org/10.1080/17470211003611264

Ki, J. J., Kelly, S. P., & Parra, L. C. (2016). Attention Strongly Modulates Reliability of Neural Responses to Naturalistic Narrative Stimuli. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *36*(10), 3092–3101. https://doi.org/10.1523/JNEUROSCI.2942-15.2016

Klein, S. B., & Kihlstrom, J. F. (1986). Elaboration, organization, and the self-reference effect in memory. *Journal of Experimental Psychology: General*, *115*(1), 26–38. https://doi.org/10.1037/0096-3445.115.1.26

Klein, S. B., & Loftus, J. (1988). The nature of self-referent encoding: The contributions of elaborative and organizational processes. *Journal of Personality and Social Psychology*, *55*(1), 5–11. https://doi.org/10.1037/0022-3514.55.1.5

Kriegeskorte, N. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*. https://doi.org/10.3389/neuro.06.004.2008

Kriegeskorte, N. (2011). Pattern-information analysis: From stimulus decoding to computational-model testing. *NeuroImage*, *56*(2), 411–421. https://doi.org/10.1016/j.neuroimage.2011.01.061

Kriegeskorte, N., Formisano, E., Sorger, B., & Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences*, *104*(51), 20600–20605. https://doi.org/10.1073/pnas.0705654104

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868. https://doi.org/10.1073/pnas.0600244103

Krienen, F. M., Tu, P.-C. P.-C., & Buckner, R. L. (2010). Clan mentality: Evidence that the medial prefrontal cortex responds to close others. *The Journal of Neuroscience*, *30*(41), 13906–13915. https://doi.org/10.1523/JNEUROSCI.2180-10.2010

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*(1), 22–44. https://doi.org/10.1037/0033-295X.99.1.22

Kuiper, N. A., & Rogers, T. B. (1979). Encoding of personal information: Self–other differences. *Journal of Personality and Social Psychology*, *37*(4), 499–514. https://doi.org/10.1037/0022-3514.37.4.499

Kumar, M., Anderson, M. J., Antony, J. W., Baldassano, C., Brooks, P. P., Cai, M. B., Chen, P.-H. C., Ellis, C. T., Henselman-Petrusek, G., Huberdeau, D., Hutchinson, J. B., Li, Y. P., Lu, Q., Manning, J. R., Mennen, A. C., Nastase, S. A., Richard, H., Schapiro, A. C., Schuck, N. W., … Norman, K. A. (2021). BrainIAK: The Brain Imaging Analysis Kit. *Aperture Neuro*, *1*, 1–19. https://doi.org/10.52294/31bb5b68-2184-411b-8c00-a1dacb61e1da

Kumaran, D., Banino, A., Blundell, C., Hassabis, D., Dayan, P., Adolphs, R., Amodio, D. M., Frith, C. D., Apps, M. A. J., Rushworth, M. F. S., Chang, S. W. C., Ashby, F. G., Waldschmidt, J. G., Beckmann, M., Johansen-Berg, H., Rushworth, M. F. S., Behrens, T. E. J., Hunt, L. T., Rushworth, M. F. S., … Meyer-Lindenberg, A. (2016). Computations Underlying Social Hierarchy Learning: Distinct Neural Mechanisms for Updating and

Representing Self-Relevant Information. *Neuron*, *92*(5), 1135–1147. https://doi.org/10.1016/j.neuron.2016.10.052

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review*, *103*(2), 284–308. https://doi.org/10.1037/0033-295X.103.2.284

Kurth, F., Zilles, K., Fox, P. T., Laird, A. R., & Eickhoff, S. B. (2010). A link between the systems: Functional differentiation and integration within the human insula revealed by meta-analysis. *Brain Structure & Function*, *214*(5–6), 519–534. https://doi.org/10.1007/s00429-010-0255-z

Lau, T., Gershman, S. J., & Cikara, M. (2020). Social structure learning in human anterior insula. *eLife*, *9*, e53162. https://doi.org/10.7554/eLife.53162

Lebreton, M., Jorge, S., Michel, V., Thirion, B., & Pessiglione, M. (2009). An automatic valuation system in the human brain: Evidence from functional neuroimaging. *Neuron*, *64*(3), 431–439. https://doi.org/10.1016/j.neuron.2009.09.040

Leech, R., & Sharp, D. J. (2013). The role of the posterior cingulate cortex in cognition and disease. *Brain : A Journal of Neurology*. https://doi.org/10.1093/brain/awt162

Leopold, A., Krueger, F., dal Monte, O., Pardini, M., Pulaski, S. J., Solomon, J., & Grafman, J. (2012). Damage to the left ventromedial prefrontal cortex impacts affective theory of mind. *Social Cognitive and Affective Neuroscience*, *7*(8), 871–880. https://doi.org/10.1093/scan/nsr071

Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic Mapping of a Hierarchy of Temporal Receptive Windows Using a Narrated Story. *Journal of Neuroscience*, *31*(8), 2906–2915. https://doi.org/10.1523/JNEUROSCI.3684-10.2011

Levorsen, M., Aoki, R., Matsumoto, K., Sedikides, C., & Izuma, K. (2023). The self-concept is represented in the medial prefrontal cortex in terms of self-importance. *The Journal of Neuroscience*, JN-RM-2178-22. https://doi.org/10.1523/JNEUROSCI.2178-22.2023

Levy, D. J., & Glimcher, P. W. (2012). The root of all value: A neural common currency for choice. *Current Opinion in Neurobiology*, *22*(6), 1027–1038. https://doi.org/10.1016/j.conb.2012.06.001

Lim, S.-L., O'Doherty, J. P., & Rangel, A. (2011). The Decision Value Computations in the vmPFC and Striatum Use a Relative Value Code That is Guided by Visual Attention. *The Journal of Neuroscience*, *31*(37), 13214–13223. https://doi.org/10.1523/JNEUROSCI.1246-11.2011

Locke, J. (2000). *An essay concerning human understanding* (P. H. Nidditch, Ed.; Repr., 1. issued as a paperback, 16. [impr.]). Clarendon Press.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*(2), 309–332. https://doi.org/10.1037/0033-295X.111.2.309

Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *Journal of Neurophysiology*, *77*(1), 24–42. https://doi.org/10.1152/jn.1997.77.1.24

Ma, N., Baetens, K., Vandekerckhove, M., Kestemont, J., Fias, W., & Van Overwalle, F. (2014). Traits are represented in the medial prefrontal cortex: An fMRI adaptation study. *Social Cognitive and Affective Neuroscience*, *9*(8), 1185–1192. https://doi.org/10.1093/scan/nst098

Ma, N., Baetens, K., Vandekerckhove, M., Van der Cruyssen, L., & Van Overwalle, F. (2014a). Dissociation of a trait and a valence representation in the mPFC. *Social Cognitive and Affective Neuroscience*, *9*(10), 1506–1514. https://doi.org/10.1093/scan/nst143

Ma, N., Baetens, K., Vandekerckhove, M., Van der Cruyssen, L., & Van Overwalle, F. (2014b). Dissociation of a trait and a valence representation in the mPFC. *Social Cognitive and Affective Neuroscience*, *9*(10), 1506–1514. https://doi.org/10.1093/scan/nst143

Mack, A., & Rock, I. (1998). *Inattentional blindness* (pp. xiv, 273). The MIT Press.

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, *51*, 93–120. https://doi.org/10.1146/annurev.psych.51.1.93

Macrae, C. N., Moran, J. M., Heatherton, T. F., Banfield, J. F., & Kelley, W. M. (2004). Medial prefrontal activity predicts memory for self. *Cerebral Cortex*, *14*(6), 647–654. https://doi.org/10.1093/cercor/bhh025

Mak, L. E., Minuzzi, L., MacQueen, G., Hall, G., Kennedy, S. H., & Milev, R. (2017). The default mode network in healthy individuals: A systematic review and meta-analysis. *Brain Connectivity*, *7*(1), 25–33.

Markus, H. (1977). Self-schemata and processing information about the self. *Journal of Personality and Social Psychology*, *35*(2), 63–78. https://doi.org/10.1037/0022-3514.35.2.63

Markus, H., Crane, M., Bernstein, S., & Siladi, M. (1982). Self-schemas and gender. *Journal of Personality and Social Psychology*, *42*(1), 38–50. https://doi.org/10.1037/0022-3514.42.1.38

Markus, H., Hamill, R., & Sentis, K. P. (1987). Thinking fat: Self-schemas for body weight and the processing of weight relevant information. *Journal of Applied Social Psychology*, *17*(1), 50–71. https://doi.org/10.1111/j.1559-1816.1987.tb00292.x

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*(2), 224–253. https://doi.org/10.1037/0033-295X.98.2.224

Markus, H., Smith, J., & Moreland, R. L. (1985). Role of the self-concept in the perception of others. *Journal of Personality and Social Psychology*, *49*(6), 1494–1512. https://doi.org/10.1037/0022-3514.49.6.1494

Marquine, M. J., Grilli, M. D., Rapcsak, S. Z., Kaszniak, A. W., Ryan, L., Walther, K., & Glisky, E. L. (2016). Impaired personal trait knowledge, but spared other-person trait knowledge, in an individual with bilateral damage to the medial prefrontal cortex. *Neuropsychologia*, *89*, 245–253. https://doi.org/10.1016/j.neuropsychologia.2016.06.021

Mars, R. B., Neubert, F.-X., Noonan, M. P., Sallet, J., Toni, I., & Rushworth, M. F. S. (2012). On the relationship between the "default mode network" and the "social brain." *Frontiers in Human Neuroscience*, *6*, 189. https://doi.org/10.3389/fnhum.2012.00189

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, *50*(4), 370–396. https://doi.org/10.1037/h0054346

Maus, B., van Breukelen, G. J. P., Goebel, R., & Berger, M. P. F. (2010). Optimization of Blocked Designs in fMRI Studies. *Psychometrika*, *75*(2), 373–390. https://doi.org/10.1007/s11336-010-9159-3

McAdams, D. P. (1996). Personality, Modernity, and the Storied Self: A Contemporary Framework for Studying Persons. *Psychological Inquiry*, *7*(4), 295–321. https://doi.org/10.1207/s15327965pli0704_1

McAdams, D. P. (2001). The Psychology of Life Stories. *Review of General Psychology*, *5*(2), 100–122. https://doi.org/10.1037/1089-2680.5.2.100

McAdams, D. P. (2018). Narrative Identity: What Is It? What Does It Do? How Do You Measure It? *Imagination, Cognition and Personality*, *37*(3), 359–372. https://doi.org/10.1177/0276236618756704

Mead, G. H. (1934). *Mind, Self, and Society from the Standpoint of a Social Behaviorist*. Chicago.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*(3), 207–238. https://doi.org/10.1037/0033-295X.85.3.207

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: A network model of insula function. *Brain Structure & Function*, *214*(5–6), 655–667. https://doi.org/10.1007/s00429-010-0262-0

Meyer, M. L., & Lieberman, M. D. (2018). Why People Are Always Thinking about Themselves: Medial Prefrontal Cortex Activity during Rest Primes Self-referential Processing. *Journal of Cognitive Neuroscience*, *30*(5), 714–721. https://doi.org/10.1162/jocn_a_01232

Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *27*(3), 775–799.

Mitchell, J. P. (2004). Encoding-Specific Effects of Social Cognition on the Neural Correlates of Subsequent Memory. *Journal of Neuroscience*, *24*(21), 4912–4917. https://doi.org/10.1523/JNEUROSCI.0481-04.2004

Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005a). The Link between Social Cognition and Self-referential Thought in the Medial Prefrontal Cortex. *Journal of Cognitive Neuroscience*, *17*(8), 1306–1315. https://doi.org/10.1162/0898929055002418

Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005b). The Link between Social Cognition and Self-referential Thought in the Medial Prefrontal Cortex. *Journal of Cognitive Neuroscience*, *17*(8), 1306–1315. https://doi.org/10.1162/0898929055002418

Mitchell, J. P., Cloutier, J., Banaji, M. R., & Macrae, C. N. (2006). Medial prefrontal dissociations during processing of trait diagnostic and nondiagnostic person information. *Social Cognitive and Affective Neuroscience*, *1*(1), 49–55. https://doi.org/10.1093/scan/nsl007

Mitchell, J. P., Heatherton, T. F., & Macrae, C. N. (2002). Distinct neural systems subserve person and object knowledge. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(23), 15238–15243. https://doi.org/10.1073/pnas.232395699

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable Medial Prefrontal Contributions to Judgments of Similar and Dissimilar Others. *Neuron*, *50*(4), 655–663. https://doi.org/10.1016/j.neuron.2006.03.040

Mitchell, J. P., Neil Macrae, C., & Banaji, M. R. (2005). Forming impressions of people versus inanimate objects: Social-cognitive processing in the medial prefrontal cortex. *NeuroImage*, *26*(1), 251–257. https://doi.org/10.1016/j.neuroimage.2005.01.031

Mitchell, J. P., Schirmer, J., Ames, D. L., & Gilbert, D. T. (2011). Medial prefrontal cortex predicts intertemporal choice. *Journal of Cognitive Neuroscience*, *23*(4), 857–866. https://doi.org/10.1162/jocn.2010.21479

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science (New York, N.Y.)*, *229*(4715), 782–784. https://doi.org/10.1126/science.4023713

Moray, N. (1959). Attention in Dichotic Listening: Affective Cues and the Influence of Instructions. *Quarterly Journal of Experimental Psychology*, *11*(1), 56–60. https://doi.org/10.1080/17470215908416289

Musslick, S., Saxe, A., Hoskin, A. N., Sagiv, Y., Reichman, D., Petri, G., & Cohen, J. D. (2020). *On the Rational Boundedness of Cognitive Control: Shared Versus Separated Representations*. https://doi.org/10.31234/osf.io/jkhdf

Nastase, S. A., Gazzola, V., Hasson, U., & Keysers, C. (2019). Measuring shared responses across subjects using intersubject correlation. *Social Cognitive and Affective Neuroscience*, nsz037. https://doi.org/10.1093/scan/nsz037

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

Northoff, G., & Hayes, D. J. (2011). Is Our Self Nothing but Reward? *Biological Psychiatry*, *69*(11), 1019–1025. https://doi.org/10.1016/j.biopsych.2010.12.014

Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain–a meta-analysis of imaging studies on the self. *NeuroImage*, *31*(1), 440–457. https://doi.org/10.1016/j.neuroimage.2005.12.002

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology. General*, *115*(1), 39–61. https://doi.org/10.1037//0096-3445.115.1.39

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review*, *7*(3), 375–402.

Nugiel, T., & Beer, J. S. (2020). How Does Motivation Modulate the Operation of the Mentalizing Network in Person Evaluation? *Journal of Cognitive Neuroscience*, *32*(4), 664–673. https://doi.org/10.1162/jocn_a_01501

Padoa-Schioppa, C., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature*, *441*(7090), 223–226. https://doi.org/10.1038/nature04676

Parkinson, C., Kleinbaum, A. M., & Wheatley, T. (2018). Similar neural responses predict friendship. *Nature Communications*, *9*(1), 332. https://doi.org/10.1038/s41467-017-02722-7

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Perini, I., Gustafsson, P. A., Hamilton, J. P., Kämpe, R., Zetterqvist, M., & Heilig, M. (2018). The salience of self, not social pain, is encoded by dorsal anterior cingulate and insula. *Scientific Reports*, *8*(1), 6165. https://doi.org/10.1038/s41598-018-24658-8

Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: From a "low road" to "many roads" of evaluating biological significance. *Nature Reviews. Neuroscience*, *11*(11), 773–783. https://doi.org/10.1038/nrn2920

Petkov, C. I., Kang, X., Alho, K., Bertrand, O., Yund, E. W., & Woods, D. L. (2004). Attentional modulation of human auditory cortex. *Nature Neuroscience*, *7*(6), 658–663. https://doi.org/10.1038/nn1256

Pfeifer, J. H., & Berkman, E. T. (2018). The Development of Self and Identity in Adolescence: Neural Evidence and Implications for a Value-Based Choice Perspective on Motivated Behavior. *Child Development Perspectives*, *12*(3), 158–164. https://doi.org/10.1111/cdep.12279

Pfeifer, J. H., Kahn, L. E., Merchant, J. S., Peake, S. J., Veroude, K., Masten, C. L., Lieberman, M. D., Mazziotta, J. C., & Dapretto, M. (2013). Longitudinal Change in the Neural Bases of Adolescent Social Self-Evaluations: Effects of Age and Pubertal Development. *The Journal of Neuroscience*, *33*(17), 7415–7419. https://doi.org/10.1523/JNEUROSCI.4074-12.2013

Pfeifer, J. H., Lieberman, M. D., & Dapretto, M. (2007). "I Know You Are But What Am I?!": Neural Bases of Self- and Social Knowledge Retrieval in Children and Adults. *Journal of Cognitive Neuroscience*, *19*(8), 1323. https://doi.org/10.1162/jocn.2007.19.8.1323

Pfeifer, J. H., Masten, C. L., Borofsky, L. A., Dapretto, M., Fuligni, A. J., & Lieberman, M. D. (2009). Neural correlates of direct and reflected self-appraisals in adolescents and adults: When social perspective-taking informs self-perception. *Child Development*, *80*(4), 1016–1038.

Pfeifer, J. H., & Peake, S. J. (2012). Self-development: Integrating cognitive, socioemotional, and neuroimaging perspectives. *Developmental Cognitive Neuroscience*, *2*(1), 55. https://doi.org/10.1016/j.dcn.2011.07.012

Philiastides, M. G., Biele, G., & Heekeren, H. R. (2010). A mechanistic account of value computation in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(20), 9430–9435. https://doi.org/10.1073/pnas.1001732107

Piaget, J. (1952). *The origins of intelligence in children* (p. 419). W W Norton & Co. https://doi.org/10.1037/11494-000

Plato, & Segal, E. (1986). *The dialogues of Plato*. Bantam Books.

Posner, M. I., & Dehaene, S. (1994). Attentional networks. *Trends in Neurosciences*, *17*(2), 75–79. https://doi.org/10.1016/0166-2236(94)90078-7

Preti, M. G., Bolton, T. A., & Van De Ville, D. (2017). The dynamic functional connectome: State-of-the-art and perspectives. *NeuroImage*, *160*, 41–54. https://doi.org/10.1016/j.neuroimage.2016.12.061

Puusepp, V. (2023). Becoming closer to one another: Shared emotions and social relationships. *Philosophical Psychology*, *0*(0), 1–27. https://doi.org/10.1080/09515089.2023.2171858

Qin, P., & Northoff, G. (2011). How is our self related to midline regions and the default-mode network? *NeuroImage*, *57*(3), 1221–1233. https://doi.org/10.1016/j.neuroimage.2011.05.028

Quadflieg, S., Flannigan, N., Waiter, G. D., Rossion, B., Wig, G. S., Turk, D. J., & Macrae, C. N. (2011). Stereotype-based modulation of person perception. *NeuroImage*, *57*(2), 549–557. https://doi.org/10.1016/j.neuroimage.2011.05.004

Raichle, M. E. (2015). The Brain's Default Mode Network. *Annual Review of Neuroscience*, *April*, 413–427.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of*

Sciences of the United States of America, *98*(2), 676–682. https://doi.org/10.1073/pnas.98.2.676

Rameson, L. T., Satpute, A. B., & Lieberman, M. D. (2010). The neural correlates of implicit and explicit self-relevant processing. *NeuroImage*, *50*(2), 701–708. https://doi.org/10.1016/j.neuroimage.2009.12.098

Ranganath, C., & Ritchey, M. (2012). Two cortical systems for memory-guided behaviour. *Nature Reviews Neuroscience*, *13*(10), 713–726. https://doi.org/10.1038/nrn3338

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

Rangel, A., & Clithero, J. A. (2014). Chapter 8—The Computation of Stimulus Values in Simple Choice. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics (Second Edition)* (pp. 125–148). Academic Press. https://doi.org/10.1016/B978-0-12-416008-8.00008-5

Raykov, P. P., Keidel, J. L., Oakhill, J., & Bird, C. M. (2020). The brain regions supporting schema-related processing of people's identities. *Cognitive Neuropsychology*, *37*(1–2), 8–24. https://doi.org/10.1080/02643294.2019.1685958

Raykov, P. P., Keidel, J. L., Oakhill, J., & Bird, C. M. (2021). Activation of Person Knowledge in Medial Prefrontal Cortex during the Encoding of New Lifelike Events. *Cerebral Cortex*, *31*(7), 3494–3505. https://doi.org/10.1093/cercor/bhab027

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, *3*(3), 382–407. https://doi.org/10.1016/0010-0285(72)90014-X

Regev, M., Simony, E., Lee, K., Tan, K. M., Chen, J., & Hasson, U. (2019). Propagation of Information Along the Cortical Hierarchy as a Function of Attention While Reading and

Listening to Stories. *Cerebral Cortex*, *29*(10), 4017–4034. https://doi.org/10.1093/cercor/bhy282

Robinson, A. K., Rich, A. N., & Woolgar, A. (2022). Linking the Brain with Behavior: The Neural Dynamics of Success and Failure in Goal-directed Behavior. *Journal of Cognitive Neuroscience*, *34*(4), 639–654. https://doi.org/10.1162/jocn_a_01818

Rogers, K. H. (2021). The Role of Normative Information in Judgments of Others. In T. D. Letzring & J. S. Spain (Eds.), *The Oxford Handbook of Accurate Personality Judgment* (pp. 234–244). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190912529.013.15

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, *35*(9), 677–688. https://doi.org/10.1037/0022-3514.35.9.677

Rosch, E. (2002). *Principles of categorization* (p. 270). MIT Press.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573–605. https://doi.org/10.1016/0010-0285(75)90024-9

Roy, M., Shohamy, D., & Wager, T. D. (2012). Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends in Cognitive Sciences*, *16*(3), 147–156. https://doi.org/10.1016/j.tics.2012.01.005

Rudebeck, P. H., Bannerman, D. M., & Rushworth, M. F. S. (2008). The contribution of distinct subregions of the ventromedial frontal cortex to emotion, social behavior, and decision making. *Cognitive, Affective & Behavioral Neuroscience*, *8*(4), 485–497. https://doi.org/10.3758/CABN.8.4.485

Santiesteban, I., Banissy, M. J., Catmur, C., & Bird, G. (2012). Enhancing Social Ability by Stimulating Right Temporoparietal Junction. *Current Biology*, *22*(23), 2274–2277. https://doi.org/10.1016/j.cub.2012.10.018

Sartre, J.-P., & Sartre, J.-P. (2012). *Being and nothingness: An essay on phenomenological ontology* (23rd print). Washington Square Press.

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, *16*(2), 235–239. https://doi.org/10.1016/j.conb.2006.03.001

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking peopleThe role of the temporo-parietal junction in "theory of mind." *NeuroImage*, *19*(4), 1835–1842.

Saxe, R. R., Whitfield-Gabrieli, S., Scholz, J., & Pelphrey, K. A. (2009). Brain Regions for Perceiving and Reasoning About Other People in School-Aged Children. *Child Development*, *80*(4), 1197–1209. https://doi.org/10.1111/j.1467-8624.2009.01325.x

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: The prospective brain. *Nature Reviews. Neuroscience*, *8*(9), 657–661. https://doi.org/10.1038/nrn2213

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. T. (2018). Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral Cortex*, *28*(9), 3095–3114. https://doi.org/10.1093/cercor/bhx179

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical*

Transactions of the Royal Society B: Biological Sciences, *372*(1711), 20160049. https://doi.org/10.1098/rstb.2016.0049

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.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *The Behavioral and Brain Sciences*, *36*(4), 393–414. https://doi.org/10.1017/S0140525X12000660

Schiller, D., Freeman, J. B., Mitchell, J. P., Uleman, J. S., & Phelps, E. a. (2009). A neural mechanism of first impressions. *Nature Neuroscience*, *12*(4), 508–514. https://doi.org/10.1038/nn.2278

Schlichting, M. L., & Preston, A. R. (2016). Hippocampal-medial prefrontal circuit supports memory updating during learning and post-encoding rest. *Neurobiology of Learning and Memory*, *134 Pt A*(Pt A), 91–106. https://doi.org/10.1016/j.nlm.2015.11.005

Schmitz, T. W., & Johnson, S. C. (2007). Relevance to self: A brief review and framework of neural systems underlying appraisal. *Neuroscience & Biobehavioral Reviews*, *31*(4), 585–596. https://doi.org/10.1016/j.neubiorev.2006.12.003

Schneider, B., & Koenigs, M. (2017). Human lesion studies of ventromedial prefrontal cortex. *Neuropsychologia*, *107*, 84–93. https://doi.org/10.1016/j.neuropsychologia.2017.09.035

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, *80*(1), 1–46. https://doi.org/10.1016/S0010-0277(00)00152-9

Schüller, A., Schilling, A., Krauss, P., Rampp, S., & Reichenbach, T. (2023). Attentional Modulation of the Cortical Contribution to the Frequency-Following Response Evoked

by Continuous Speech. *Journal of Neuroscience*, *43*(44), 7429–7440. https://doi.org/10.1523/JNEUROSCI.1247-23.2023

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*, *147*(3), 293–327. https://doi.org/10.1037/bul0000303

Schuwerk, T., Schurz, M., Müller, F., Rupprecht, R., & Sommer, M. (2017). The rTPJ's overarching cognitive function in networks for attention and theory of mind. *Social Cognitive and Affective Neuroscience*, *12*(1), 157–168. https://doi.org/10.1093/scan/nsw163

Sebastian, C., Burnett, S., & Blakemore, S.-J. (2008). Development of the self-concept during adolescence. *Trends in Cognitive Sciences*, *12*(11), 441–446.

Shapiro, K. L., Raymond, J. E., & Arnell, K. M. (1997). The attentional blink. *Trends in Cognitive Sciences*, *1*(8), 291–296. https://doi.org/10.1016/S1364-6613(97)01094-2

Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences*, *111*(43), E4687–E4696. https://doi.org/10.1073/pnas.1323812111

Simony, E., Honey, C. J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., & Hasson, U. (2016). Dynamic reconfiguration of the default mode network during narrative comprehension. *Nature Communications*, *7*(1), 12141. https://doi.org/10.1038/ncomms12141

Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering: Empirically navigating the stream of consciousness. *Annual Review of Psychology*, *66*, 487–518. https://doi.org/10.1146/annurev-psych-010814-015331

Smirnov, D., Glerean, E., Lahnakoski, J. M., Salmi, J., Jääskeläinen, I. P., Sams, M., & Nummenmaa, L. (2014). Fronto-parietal network supports context-dependent speech comprehension. *Neuropsychologia*, *63*, 293–303. https://doi.org/10.1016/j.neuropsychologia.2014.09.007

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1411–1436. https://doi.org/10.1037/0278-7393.24.6.1411

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., & Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, *23*, S208–S219. https://doi.org/10.1016/j.neuroimage.2004.07.051

Spalding, K. N., Schlichting, M. L., Zeithamova, D., Preston, A. R., Tranel, D., Duff, M. C., & Warren, D. E. (2018). Ventromedial Prefrontal Cortex Is Necessary for Normal Associative Inference and Memory Integration. *The Journal of Neuroscience*, *38*(15), 3767–3775. https://doi.org/10.1523/JNEUROSCI.2501-17.2018

Spreng, R. N., & Grady, C. L. (2010). Patterns of brain activity supporting autobiographical memory, prospection, and theory of mind, and their relationship to the default mode network. *Journal of Cognitive Neuroscience*, *22*(6), 1112–1123. https://doi.org/10.1162/jocn.2009.21282

Spreng, R. N., Mar, R. A., & Kim, A. S. N. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: A quantitative meta-analysis. *Journal of Cognitive Neuroscience*, *21*(3), 489–510. https://doi.org/10.1162/jocn.2008.21029

Spreng, R. N., Stevens, W. D., Chamberlain, J. P., Gilmore, A. W., & Schacter, D. L. (2010). Default network activity, coupled with the frontoparietal control network, supports goal-directed cognition. *NeuroImage*, *53*(1), 303–317. https://doi.org/10.1016/j.neuroimage.2010.06.016

Stolier, R. M., & Freeman, J. B. (2016). Neural pattern similarity reveals the inherent intersection of social categories. *Nature Neuroscience*, *19*(6), 795–797. https://doi.org/10.1038/nn.4296

Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *The Behavioral and Brain Sciences*, *30*(3), 299–313; discussion 313-351. https://doi.org/10.1017/S0140525X07001975

Sui, J., He, X., & Humphreys, G. W. (2012). Perceptual effects of social salience: Evidence from self-prioritization effects on perceptual matching. *Journal of Experimental Psychology. Human Perception and Performance*, *38*(5), 1105–1117. https://doi.org/10.1037/a0029792

Sui, J., & Humphreys, G. W. (2015a). The Integrative Self: How Self-Reference Integrates Perception and Memory. *Trends in Cognitive Sciences*, *19*(12), 719–728. https://doi.org/10.1016/j.tics.2015.08.015

Sui, J., & Humphreys, G. W. (2015b). The Integrative Self: How Self-Reference Integrates Perception and Memory. *Trends in Cognitive Sciences*, *19*(12), 719–728. https://doi.org/10.1016/j.tics.2015.08.015

Sui, J., & Rotshtein, P. (2019). Self-prioritization and the attentional systems. *Current Opinion in Psychology*, *29*, 148–152. https://doi.org/10.1016/j.copsyc.2019.02.010

Sui, J., Rotshtein, P., & Humphreys, G. W. (2013). Coupling social attention to the self forms a network for personal significance. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(19), 7607–7612. https://doi.org/10.1073/pnas.1221862110

Sui, J., Zhu, Y., & Han, S. (2006). Self-face recognition in attended and unattended conditions: An event-related brain potential study. *NeuroReport*, *17*(4), 423. https://doi.org/10.1097/01.wnr.0000203357.65190.61

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403–409. https://doi.org/10.1016/j.tics.2009.06.003

Svoboda, E., McKinnon, M. C., & Levine, B. (2006). The functional neuroanatomy of autobiographical memory: A meta-analysis. *Neuropsychologia*, *44*(12), 2189–2208. https://doi.org/10.1016/j.neuropsychologia.2006.05.023

Swencionis, J. K., & Fiske, S. T. (2013). More human: Individuation in the 21st century. In *In Humanness and dehumanization* (pp. 276–293). Psychology Press.

Syed, M., DeYoung, C. G., & Tiberius, V. (2019). Self, Motivation, and Virtue, or How We Learned to Stop Worrying and Love Deep Integration. *Self, Motivation, and Virtue*, *48*, 7–24. https://doi.org/10.4324/9780429260858-2

Symons, C. S., & Johnson, B. T. (1997). The self-reference effect in memory: A meta-analysis. *Psychological Bulletin*, *121*(3), 371–394. https://doi.org/10.1037/0033-2909.121.3.371

Tajfel, H. (1970). Experiments in Intergroup Discrimination. *Scientific American*, *223*(5), 96–103.

Tamir, D. I., & Mitchell, J. P. (2010). Neural correlates of anchoring-and-adjustment during mentalizing. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(24), 10827–10832. https://doi.org/10.1073/pnas.1003242107

Tamir, D. I., & Thornton, M. A. (2018). Modeling the Predictive Social Mind. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2017.12.005

Tavares, R. M., Mendelsohn, A., Grossman, Y., Williams, C. H., Shapiro, M., Trope, Y., & Schiller, D. (2015). A Map for Social Navigation in the Human Brain. *Neuron*, *87*(1), 231–243. https://doi.org/10.1016/j.neuron.2015.06.011

Taylor, C. (1989). *Sources of the self: The making of the modern identity*. Harvard University Press.

Thornton, M. A., & Mitchell, J. P. (2017). Consistent Neural Activity Patterns Represent Personally Familiar People. *Journal of Cognitive Neuroscience*, *29*(9), 1583–1594. https://doi.org/10.1162/jocn_a_01151

Thornton, M. A., & Mitchell, J. P. (2018). Theories of Person Perception Predict Patterns of Neural Activity During Mentalizing. *Cerebral Cortex*, *28*(10), 3505–3520. https://doi.org/10.1093/cercor/bhx216

Thornton, M. A., Weaverdyck, M. E., & Tamir, D. I. (2019). The brain represents people as the mental states they habitually experience. *Nature Communications*, *10*(1), 2291. https://doi.org/10.1038/s41467-019-10309-7

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Tsantani, M., Kriegeskorte, N., McGettigan, C., & Garrido, L. (2019). Faces and voices in the brain: A modality-general person-identity representation in superior temporal sulcus. *NeuroImage*, *201*, 116004. https://doi.org/10.1016/j.neuroimage.2019.07.017

Uddin, L. Q. (2015). Salience processing and insular cortical function and dysfunction. *Nature Reviews Neuroscience*, *16*(1), 55–61. https://doi.org/10.1038/nrn3857

Van Dijk, K. R. A., Hedden, T., Venkataraman, A., Evans, K. C., Lazar, S. W., & Buckner, R. L. (2010). Intrinsic functional connectivity as a tool for human connectomics: Theory, properties, and optimization. *Journal of Neurophysiology*, *103*(1), 297–321. https://doi.org/10.1152/jn.00783.2009

van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, *35*(4), 211–219. https://doi.org/10.1016/j.tins.2012.02.001

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

Vatansever, D., Menon, D. K., Manktelow, A. E., Sahakian, B. J., & Stamatakis, E. A. (2015). Default Mode Dynamics for Global Functional Integration. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*(46), 15254–15262. https://doi.org/10.1523/JNEUROSCI.2135-15.2015

Vodrahalli, K., Chen, P.-H., Liang, Y., Baldassano, C., Chen, J., Yong, E., Honey, C., Hasson, U., Ramadge, P., Norman, K. A., & Arora, S. (2018). Mapping between fMRI responses to movies and their natural language annotations. *NeuroImage*, *180*, 223–231. https://doi.org/10.1016/j.neuroimage.2017.06.042

Wagner, D. D., Chavez, R. S., & Broom, T. W. (2019). Decoding the neural representation of self and person knowledge with multivariate pattern analysis and data-driven approaches. *Wiley Interdisciplinary Reviews: Cognitive Science*, *10*(1), e1482. https://doi.org/10.1002/wcs.1482

Wagner, D. D., Haxby, J. V., & Heatherton, T. F. (2012). The representation of self and person knowledge in the medial prefrontal cortex: The representation of self and person knowledge. *Wiley Interdisciplinary Reviews: Cognitive Science*, *3*(4), 451–470. https://doi.org/10.1002/wcs.1183

Welborn, B. L., & Lieberman, M. D. (2015). Person-specific Theory of Mind in Medial pFC. *Journal of Cognitive Neuroscience*, *27*(1), 1–12. https://doi.org/10.1162/jocn_a_00700

Whalen, P. J. (2007). The uncertainty of it all. *Trends in Cognitive Sciences*, *11*(12), 499–500. https://doi.org/10.1016/j.tics.2007.08.016

Whitfield-Gabrieli, S., Moran, J. M., Nieto-Castañón, A., Triantafyllou, C., Saxe, R., & Gabrieli, J. D. E. (2011). Associations and dissociations between default and self-reference networks in the human brain. *NeuroImage*, *55*(1), 225–232. https://doi.org/10.1016/j.neuroimage.2010.11.048

Wiener, N. (2007). *Cybernetics or control and communication in the animal and the machine* (2. ed., reprint). MIT Press.

Woldorff, M. G., & Hillyard, S. A. (1991). Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalography and Clinical Neurophysiology*, *79*(3), 170–191. https://doi.org/10.1016/0013-4694(91)90136-r

Yankouskaya, A., Humphreys, G., Stolte, M., Stokes, M., Moradi, Z., & Sui, J. (2017). An anterior–posterior axis within the ventromedial prefrontal cortex separates self and reward. *Social Cognitive and Affective Neuroscience*, *12*(12), 1859–1868. https://doi.org/10.1093/scan/nsx112

Yeshurun, Y., Nguyen, M., & Hasson, U. (2021). The default mode network: Where the idiosyncratic self meets the shared social world. *Nature Reviews Neuroscience*, 1–12. https://doi.org/10.1038/s41583-020-00420-w

Yoon, L., Kim, K., Jung, D., & Kim, H. (2021). Roles of the MPFC and insula in impression management under social observation. *Social Cognitive and Affective Neuroscience*, *16*(5), 474–483. https://doi.org/10.1093/scan/nsab008

Zadbood, A., Chen, J., Leong, Y. C., Norman, K. A., & Hasson, U. (2017). How We Transmit Memories to Other Brains: Constructing Shared Neural Representations Via Communication. *Cerebral Cortex*, *27*(10), 4988–5000. https://doi.org/10.1093/cercor/bhx202

Zeithamova, D., Dominick, A. L., & Preston, A. R. (2012). Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron*, *75*(1), 168–179. https://doi.org/10.1016/j.neuron.2012.05.010