Title

Shared and distinct visuomotor mechanisms underlying action processing of the self and others

Permalink

https://escholarship.org/uc/item/0v77p86z

Author

Kadambi, Akila

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Shared and Distinct Visuomotor Mechanisms Underlying Action Processing of the Self and Others

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Psychology

by

Akila Kadambi

2022

ABSTRACT OF THE DISSERTATION

Shared and Distinct Visuomotor Mechanisms Underlying Action Processing of the Self and Others

by

Akila Kadambi

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2022

Professor Hongjing Lu, Chair

My dissertation aims to reveal the visuomotor mechanisms that underlie action processing of the self and others. The dissertation will shed light on domain-general computations underlying actions across the identity of others, as well as those unique to one's own action. As methods, I will incorporate behavioral, neuroimaging, and computational techniques. I will further incorporate different types of stimuli ranging from visually impoverished point-light displays of the human body to contextually rich video stimuli. Chapter 1 examines at the behavioral level, how recognition of our own actions is affected by factors beyond visual properties, related to motoric and intrinsic participant variability. Chapter 2 extends the behavioral work to brain imaging, and further measures similarities and differences in neural activity for actions of the self and others. Chapter 3 shifts from the self to the actions of others involved in a range of social interactions. This chapter will systematize the degree of social context using methods adopted from computer vision and measure the representational space of motor and social features contributing to judgements of incongruency in the interaction.

Three questions maintain throughout the chapters:

(1) What are the visuomotor mechanisms that underlie action processing of the self and others?

(2) How distinct are neural computations for action recognition of the self and others, and to what degree do they rely on visuomotor featural processing?

(3) What visuomotor features are selective and generalizable across the type of action stimuli (ranging from visually impoverished to contextually rich)?

The dissertation of Akila Kadambi is approved.

Marco Iacoboni

Kerri Johnson

Martin Monti

Hongjing Lu, Committee Chair

University of California, Los Angeles

2022

Dedicated to my family, for your support and love, and to loved ones that passed on during my PhD.

TABLE OF CONTENTS

VII. Motoric and intrinsic factors modulate self-action recognition

VIII. Specialized neural mechanisms for self-recognition from whole-body movements

IX. Visual perception of social incongruencies: How interactions go awry

  c. Experiment 1

  d. Experiment 2

XI. Appendix

LIST OF FIGURES

LIST OF TABLES

Acknowledgements

Words cannot express my gratitude for my PhD experience, my department, my mentors, my family, my friends, my collaborators, and many others throughout my life.

Pursuing a PhD taught me the value of diligence, growth, curiosity, and kindness, which I learned from incredible examples. Foremostly, I would like to acknowledge my advisor Dr. Hongjing Lu, who exhibits those qualities, and so many more. Hongjing, your patience and intelligence are unmatched. Your kindness is evident not only in our interactions, but from everyone you encounter. This dissertation would not have been possible without your support and guidance.

I would next like to acknowledge Dr. Marco Iacoboni, such a supportive mentor, intellectually curious, and kind-hearted human being. You brought an entirely new way of thinking into my eyes, and it has been invaluable to learn from your intelligence and philosophical mind. It was always so fascinating to me, how you connected something like actions to empathy and imitation. You've shaped the field in so many positive and prosocial ways, and it's been such a privilege to learn from you.

I would also like to acknowledge Drs. Martin Monti and Kerri Johnson, both wonderful members of my committee. Martin, you are a neuroimaging expert! Thank you for your patience, consideration, kindness, as well as generous funds to continue the project. You are so patient and taught me so many principles. Kerri, I remember reading some of your work prior to meeting you, and I was excited take your Social Vision course. I so admire how you pioneered the field of social vision when it was essentially non-existent. This dissertation is inspired by so much of your work.

I would also like to acknowledge my previous mentors, without whom I would never have arrived at a PhD in the first place. Dr. Luke Miller, you inspired my interest in the brain with your intelligence and dedication, which I admired as an undergraduate research assistant. Thank you for your wonderful mentorship then, and even today throughout my PhD career. Dr. Burcu Urgen, you also inspired my interest in the brain—and gave me my first brain scan! Your work on action processing is clear inspiration for much of the work in my dissertation. Dr. Ayse Saygin, thank you for welcoming me into the lab as an undergraduate and sparking my interest in the field. I am so grateful for the many opportunities I received in the lab. Dr. Laurens Krol and Dr. Thorsten Zander, I am so grateful for the wonderful research experiences in Berlin. Thank you for encouraging undergraduate researchers toward an exciting path in academia. Dr. Virgina de Sa, thank you for welcoming me into the lab as an undergraduate research assistant and teaching me many fundamentals of research and brain-computer interfaces.

None of the work in this thesis would have been possible without the support of a wonderful team of research assistants: Sophia Baia, Elinor Yeo, Jolie Wu, Kelly Nola, Tabitha Safari, Kelly Xue, Qie Xie, and so many more: you will accomplish so much, and I am excited to see what the future has it store for you! These projects relied so much on your help. Thank you also to the excellent research assistants in the Iacoboni lab. I am excited to continue our projects together.

I would further like to acknowledge the numerous collaborators, colleagues, mentors, and friends who have inspired and shaped my way of thinking throughout my research studies: Dr. Yujia Peng, Dr. Junho Lee, Dr. Micah Johnson, Dr. James Kubricht, Dr. Phil Kellman, Dr. Christine Massey, Dr. Keith Holyoak, Dr. Robert Bjork, Dr. Elizabeth Bjork, Dr. Naomi Eisenberger, Dr. Gennady Erlikhman, Dr. Nicholas Baker, Dr. Viyehni Fuchser, Dr. Daniel Posner, Dr. Chad Hazlett, Dr. Jeffrey Knotts, Dr. Angela Pham, Dr. Jeffrey Chiang, Dr. Dylan

Dickstein, Dr. Takahiro Doi, Dr. Steven Thurman, Dr. Joseph Burling, Dr. Alan Ceaser, Saskia Giebl, Mark Straccia, Carolyn Murray, Mac Xing, Sashel Haygood, Christina Harr, Katie Silaj, Julian Cheng, Daniel Virtheim, Ana Xingru Yin, Nitya Rao, Anjana Rao, Emily Garverick, Nilu Karimi, Jessica Johnson, Aneesha Grama, Shuhao Fu, Emily Wong, Alessandra Dallavecchia, Lucy Shen, Yi-Chia Chen, Qie Xie, Nicholas Ichien, Shuwen Qiu, Lisa Lee, members of the Lu lab, and so many others.

Finally, I would like to thank the most important people in my life: my family. Thank you for bringing me into this world. And for your care, kindness, and love throughout my life.

Two chapters in the thesis are either under review or published. One chapter is in preparation:

Chapter 1 is a manuscript under review with contributions from the following co-authors:

> Kadambi, A., Xie, Q., & Lu, H. (under review). Motoric and intrinsic factors modulate self-recognition of actions.

Chapter 2 is in preparation.

Chapter 3 is published in Attention, Perception, and Psychophysics, with contributions from the following co-authors:

> Kadambi, A., Ichien, N., Qiu, S., & Lu, H. (2020). Understanding the visual perception of awkward body movements: How interactions go awry. Attention, Perception, & Psychophysics, 82(5), 2544-2557.

VITA

Education

2016 - 2017	**MA**, University of California, Los Angeles ○ Cognitive Psychology, Minor in Quantitative Psychology
2011- 2015	**BS**, University of California, San Diego ○ Cognitive Science, with emphasis in Computation, Minor in Psychology

Publications

Kadambi, A., Ichien, N., Qiu, S. & Lu, H. (2020). Understanding the visual perception of awkwardness: How greetings go awry. Attention, Perception, & Psychophysics, 1-14.

Featured article (Psychonomics): link

Kadambi, A & Lu, H. (2019). Individual differences in self-recognition from body movements. Proceedings of the 41st Annual Meeting of the Cognitive Science Society. Montreal, Canada: Cognitive Science Society.

Burling, J. M., **Kadambi, A.,** Safari, T., & Lu, H. (2019). The impact of autistic traits on self-recognition of body movements. Frontiers in Psychology, 9, 2687.

Manuscripts Under Review

Kadambi, A., Qie, X, & Lu, H. (under review). Motoric and intrinsic factors modulate self-recognition from actions.

Conference Talks and Posters

Baia, S., **Kadambi, A.,** Lu, H., *Contextual and Kinematic Contributions to Expertise Identification in Yoga.* Vision Science Society, Virtual Conference (May 2021)

Kadambi, A., Erlikhman, G., Johnson, M., Monti, M., Lu, H. *Brain networks for visual self-recognition from whole-body movements* (Jan 2020). Society for Neuroscience (SFN), Virtual Conference$^+$

Kadambi, A., Erlikhman, G., Monti, M., Lu, H. *Brain networks for visual self-recognition from whole-body movements* (May 2020). Vision Science Society (VSS), Virtual Conference$^+$

Kadambi, A., Lu, H. *Individual Differences in Self-Recognition from Body Movements* (July 2019). Cognitive Sciences Society in Montreal, Canada*$^+$

Kadambi, A., Lu, H., *Social Threat Perception from Body Movements* (May 2019). Vision Science Society (VSS) in St. Pete's Beach, Florida.

Kadambi, A., Lu, H., *Individual Differences in Self-recognition from Body Movements* **(**May 2018). Vision Science Society (VSS) in St. Pete's Beach, Florida.

Kadambi, A., Ichien, N., Qiu, S., Lu, H., *That was awkward! How greetings go awry* (May 2018). Vision Science Society (VSS) in St. Pete's Beach, Florida

Kadambi, A., Pham, A., Miller, L., & Saygin, A. (May 2015). *Biological motion processing under interocular suppression.* Vision Sciences Society (VSS) in St. Pete's Beach, Florida

Kadambi, A., Pham, A., Miller, L., & Saygin, A. (June 2014). *Biological motion processing under interocular suppression.* UCSD Summer Research Conference, La Jolla, CA.

* *invited talk*

$^+$ *award recipient*

Kadambi, A., Miller, L., & Saygin, A. (June 2014). *Unconscious processing of biological motion.* UCSD Summer Research Conference, La Jolla, CA.

Kadambi, A., Miller, L., & Saygin, A. (June 2013). *Unconscious processing of biological motion.* Qualcomm Institute Summer Conference, La Jolla, CA.

Awards, Grants, and Fellowships

2022	Racial Social Justice Grant*
2021	Transdisciplinary Research Acceleration Grant*
2020	APA Dissertation Research Award
2020	Society for Neuroscience Trainee Award
2020	UCLA Cognitive Research Award
2020	Elsevier/Vision Research Travel Award
2020	Bedari Kindness Institute Faculty Grant*
2019	Robert J. Glushko Travel Award (Cognitive Science Society)
2018	National Defense Science and Engineering Graduate Fellowship Semifinalist
2017	UCLA Graduate Student Summer Research Fellowship
2016	UCLA Edward W. Pauley Fellowship
2015	Distinction in Cognitive Science, UCSD
2015	Phi Beta Kappa, UCSD chapter
2014	Technische Universität Summer Research Fellow (*Berlin, Germany*)
2013	Qualcomm Institute Summer Scholar (*San Diego, CA)*
2012	Provost Honors (2012 – 2015)

** joint contributor*

Teaching

Fall 2021	Teaching Fellow: Introduction to Cognitive Psychology
Spring 2021	Teaching Fellow: MATLAB Programming for Behavioral Sciences
Winter 2021	Teaching Fellow: MATLAB Programming for Behavioral Sciences
Fall 2020	Teaching Fellow: MATLAB Programming for Behavioral Sciences
Spring 2020	Teaching Fellow: General Psychology Laboratory
Winter 2020	Teaching Associate: Introduction to Cognitive Psychology
Fall 2019	Teaching Associate: Research Methods/Experimental Design
Spring 2019	Teaching Associate: Introduction to Cognitive Psychology
Winter 2019	Teaching Associate: Cognitive Science Laboratory: Introduction to Theory and Simulation
Fall 2018	Teaching Associate: Research Methods/Experimental Design
Spring 2018	Teaching Assistant: Research Methods/Experimental Design
Winter 2018	Teaching Assistant: Psychological Statistics
Fall 2017	Teaching Assistant: Introduction to Cognitive Psychology
Fall 2013	Undergraduate Teaching Assistant: Introduction to Cognitive Science

CHAPTER 1

1.1. General Introduction

Human beings are one in the same. Our neurobiological similarities bind us together and allow for the exchange of shared experiences. Despite the similarities we share as species— our differences could collectively be reduced to what we lack access to: internal mental states of others.

Aside from verbal communication, actions are perhaps the most important medium to indirectly access the hidden mental states of others. From action information conveyed by others, humans can infer rich social attributes including identity (Jokisch et al., 2006; Loulaet al., 2005; Cutting Kozlowski, 1977; Beardsworth Buckner, 1981), familiarity (Cutting & Kozlowski, 1977), emotional state (Atkinson et al., 2004; Dittrich et al., 1996; De Gelder et al., 2015; Roether et al., 2009; Coulson, 2004) across different cultures (Parkinson et al., 2017), gender (Pollick et al., 2002; Pollick et al., 2005; Johnson et al., 2011), as well as personality traits (Thoresen et al., 2012; Heberlein et al., 2004; Gunns, Johnston, Hudson, 2002). Human inference from actions exceeds even that of modern-day surveillance systems, enacting a critical role in situations where cues to facial identity are masked (e.g., robbers covering their faces with masks) (Coste, 2020), as well as in daily life— shown recently in the pandemic, where the use of masks obscured facial identity, yet human movements proved especially useful as compensatory and robust biometric cues to identity. Despite the importance of action information in daily life, how our brain extracts high-level attributes such as identity and social inferences from sparse action information, remains largely unclear.

Findings from systems-level neuroscience demonstrate the importance of lower-level action systems to higher-level social inferences. These motoric regions are not only involved in action production and perception, but also assist in social attributions, revealed by increased neural activity within and connectivity between these regions that interact with networks typically implicated in social cognition (e.g., mentalizing network) (Keysers and Gazzola, 2007). Motoric regions that code between both the self and other are collectively termed "mirroring" regions, since they procure a special common coding function that compares the similarity between the visual processing of actions of others and our own motor codes (Di Pellegrino et al., 1992; Rizzolatti et al., 1996; Gallese et al., 1996; Rizzolatti & Craighero, 2004). From this has followed numerous studies showing how our motor experience serves as a building block for ability to engage in higher-level shared social understanding in relation to empathy and theory of mind of others (e.g., Gallese & Goldman, 1998; Gallese 2005; Carr et el., 2003; Pfeifer et al., 2008; Iacoboni 2009).

Interestingly however, we have copious visual experience observing the actions of others. It's a familiar sight to see another human being moving. It's also a familiar experience to "feel" oneself moving from a first-person and kinesthetic perspective. Unfamiliar to us, is the bridge that binds them: seeing our whole selves moving. Third-person glimpses of our bodies may indeed be captured through videos or glass mirrors, but are far less observable in daily life as compared to the rich visual experiences we have seeing the movements of close friends or coworkers. Since the visual experience of seeing oneself from an allocentric (third-person; 3PP) perspective is unfamiliar to a large extent, self-action recognition is often thought to tap into an intrinsic sense of sensorimotor or kinesthetic "awareness" (Blanke, 2012; Blanke, 2015). This fundamental difference between actions of the self and others raises important mechanistic considerations: To what extent are the neural computations of these stimuli, actions of the self and others, processed similarly in the brain? Do they rely on identical neurocomputational features as cues to identity processing? How does the presence of contextual information affect these computations? The present dissertation aims to address these questions. I will begin with a brief literature review on empirical work connecting action processing of the self and others. I will then introduce our large-scale study on self-action recognition, which in **Chapter 1**, directly compares the contribution of quantitatively derived features of visual action similarity to motoric and intrinsic influences in self-action recognition. In **Chapter 2,** I will extend the behavioral work to a neuroimaging paradigm. Here, I aim to show the commonalities and differences in neural recruitment and visuomotor features involved for action recognition of the self and others. In **Chapter 3**, I shift focus from the self, to measure the visuomotor mechanisms that underlie action processing of others. Chapter 3 incorporates advances in computer vision that systematize the relative contributions of visuomotor and contextual information. In contrast to the minimal visual input of actions in Chapters 1 and 2, Chapter 3 uses diverse and naturalistic stimuli across a range of scenes and social interactions that convey both action congruencies and incongruencies. Across these studies, findings from my thesis stand to advance our understanding of how the brain represents actions of the self and others, bridging together behavioral, neuroimaging, and computational work.

1.2. Background

1.2.1. Self-Action Processing

Since we rarely view our own body movements in our daily lives, understanding how we recognize our own movements sheds light on the core of self-awareness and on the representation of actions. The extent to which vision modulates recognition of our whole bodies in motion, posits an intriguing question. We often view our own bodies from an egocentric frame of reference, with the world constructed through these body-centered coordinates. As a result, the visual information conveyed from an egocentric perspective is incongruent with our whole bodies perceived from an allocentric frame of reference. On the other hand, we have copious experience viewing other people's actions in an allocentric frame of reference, for which we lack direct visual experience of our own whole-body actions. How does this lead to processing differences at the neural level between self and other actions? Are there unique computations when viewing our actions in the third person relative to others, even after degrading visual cues to identity?

To study the mechanisms of action processing, empirical paradigms in action recognition of the self and others often incorporate visually impoverished stimuli known as point-light displays (PLD)s (Johansson, 1973) in 3PP. These stimuli prioritize tight methodological control, by rendering visual information of the body to a series of dots localized to key joints, which enables the vivid percept of an animated human figure. Despite the visually degraded stimuli, humans can infer a wealth of socially-relevant attributes from PLDs including gender (Pollick et al., 2002; Pollick et al., 2005; Johnson et al., 2011), emotional state (Atkinson et al., 2004; Dittrich et al., 1996; De Gelder et al., 2015; Roether et al., 2009; Coulson, 2004), and identity (Cutting Kozlowski, 1977). Moreover, the work has shown compelling evidence for a self-processing advantage—that humans can recognize their own identity from PLDs presented across different action types and in these unfamiliar, allocentric perspectives (Jokisch et al., 2006; Loula et al., 2005; Beardsworth Buckner, 1981; Coste et al., 2021; Burling et al., 2019; Kadambi et al., under review). As a result, self-recognition from PLDs is often thought to rely on factors that are deeper than vision, such as motoric experience—a privileged access one has to their own motor system— since the displays are depicted in unfamiliar viewpoints and visual cues to identity are significantly diminished. However, direct evidence for the increased motoric involvement underlying the processing of self-actions is poorly understood, requiring a bridge between quantitative and neural metrics in self-action recognition. In Chapter 1, I directly test this question by measuring quantitative contributions of visual features in comparison to motoric mechanisms in self-action recognition. In Chapter 2, I extend the behavioral work to directly map out the neural circuitry that underlies self-action recognition and assess the contribution of visuomotor features in the underlying representational space.

1.2.2. Neural networks underlying action processing of self and others

Since humans can recognize their own actions from sparse visual input, other modalities and systems beyond vision clearly contribute to self-recognition and potentially to action recognition of others. What are these underlying neural mechanisms?

The neural mechanisms involved in action processing are collectively studied in two lines of work that together comprise the action observation network (AON). The first line of empirical work is focused on the visual neuroscience of actions (Urgen et al 2019), which prioritizes methodological control by using whole-body PLDs and focuses on the neural contributions of low- and mid-level temporooccipital neural regions to action processing. These studies reveal insight as to how regions of the temporooccipital cortex, notably the extrastriate body area (EBA), fusiform body area (FBA), and posterior superior temporal sulcus (pSTS), are vital for processing lower and mid-level features of the human body related to postural attributes (de Gelder and Solanas, 2021), kinematics (Astafiev et al., 2004), their spatiotemporal integration (Downing Peelen, 2001; Zimmermann et al., 2018), as well as affective attributions (e.g., De Gelder and Solanas, 2021). The EBA, in particular, sits at interface of both perceptual and motor processing, with evidence for processing both static and dynamic postural attributes related to the human body (Zimmermann et al., 2018; De Gelder and Solanas, 2021; Downing et al., 2006; Walbrin et al., 2019; Orgs et al., 2016). The pSTS is notably sensitive to biological motion processing as well as higher-level processing of the action including intentions (Saxe et al., 2004; Brass et al., 2007), valence (Verosky Todorov, 2010; Candidi et al., 2015), and social interactions (Masson Isik, 2021; Isik et al., 2017), and is further thought to be the integratory mechanism between spatiotemporal information derived from actions. Regions of the fusiform cortex in the FBA appear attuned to the holistic form of the body, where the FBA is also shown to play a role in identity processing based on bodily information (Orgs et al., 2015; de Gelder and Solanas, 2021) and prioritized processing of the whole-body relative to the EBA (Taylor et al., 2007; Hodzic et al., 2009; Orgs et al., 2016). The FBA's role has also been pinpointed to prioritize static images of the whole-body, whereas it thought that dynamic forms of the body may instead be processed by EBA (cf Vangeneugden et al., 2014) or the pSTS (Vangeneugden et al., 2014; Grossman and Blake, 2002; Grèzes et al., 2007; Grossman et al., 2010). Notably, however, the degree to which these temporooccipital regions are sensitive to the self—in particular, the self in motion— is poorly understood.

Within the temporooccipital regions, the pSTS serves as the input to the mirror neuron system (MNS) (Iacoboni et al., 2001; Rizzolati, 2005; Iacoboni & Dapretto, 2006), leading to the second line of empirical work— the cognitive neuroscience of actions (Urgen et al., 2019). From the temporooccipital regions, the flow of information processing travels anteriorly toward the frontoparietal circuitry. This circuitry that conveys input from the pSTS to the frontoparietal regions comprises the MNS (e.g., Rizzolatti & Craighero, 2004; Iacoboni & Dapretto, 2006; Iacoboni 2009). Originally discovered in area F5 in monkeys (later in other areas including the PF, area 7b), neuroimaging studies reveal considerable evidence of mirror neuron involvement at the system-level in analogous premotor, parietal, and occipital regions in humans, such as the inferior frontal gyrus (IFG), inferior parietal lobe (IPL), and the precentral gyrus (Rizzolatti et al., 1996; Buccino et al., 2001; Iacoboni et al., 1999; Rizzolatti & Craighero; 2004; Iacoboni & Dapretto, 2006). Importantly the MNS is considered the integratory system between motor production and visual observation with a degree of functional equivalence, as the regions involved in motor production also contain special visuomotor neurons that map the actions of others to one's own motor system (Rizzolatti & Craighero; 2004; Jeannerod, 2001; Decety 1994). Cognitive neuroscience approaches as studied to implicate the MNS notably differ from those in visual neuroscience due to the types of stimuli used to evoke responses in these regions, specifically relying on contextually rich video stimuli of isolated body parts (e.g., hand movements) or hand-object interactions. Regardless, the accounts together, involving the frontoparietal (cognitive neuroscience) and the temporooccipital (visual neuroscience) regions comprise the core circuitry of the action observation network (AON).

The architecture of the AON regions (temporooccipital and frontoparietal) is thought to comprise either a hierarchical structure of action processing or enact more distributed representations across the brain (Haxby et al., 2000; De Gelder & Solanas, 2021). At the lower-level of a possible processing hierarchy, one theory is that the temporooccipital regions including the pSTS, EBA, FG, inferior occipital gyrus are involved in information processing related to coarser features of the action such as shape, posture, motion patterns, or positional changes of the body. The pSTS then acts as the driving input to the MNS, where the frontoparietal regions, centered on the IPL and posterior IFG, are involved in higher-level action processing (cf Wurm and Cazzaminga, 2021), such as encoding of the action outcomes (Rizzolatti & Sinigaglia, 2010; Hamilton and Grafton, 2006) or goals (Hamilton and Grafton 2006; 2007; Grafton and Hamilton, 2007; Iacoboni et al., 2005), across effectors (Jastorff and Begliomini, 2010; Hamilton and Grafton, 2008; Chaminade et al., 2005; Ferri et al., 2015) and type of actions (Ferri et al., 2015; Fogassi et al., 2005; Goodale & Milner, 1992). Note however, it is unclear as to whether the temporooccipital regions primarily represent low/mid-level action features, such as kinematics (e.g., speed, acceleration, jerk, motion direction), postural attributes or general coherence of the stimuli, or whether they also encode higher-level conceptual understanding. There is ample evidence pointing to how the frontoparietal regions (particularly the anterior portions of the IPL) may not necessarily abstract to the level of conceptual knowledge (e.g., Goldenberg & Spatt, 2009; Wurm & Caramazza, 2021; cf Leshinskaya and Caramazza, 2015), but rather process the perceptual dependencies related to the action information, while the temporooccipital regions are responsible for early conceptual decoding. In this regard, (also known as) cognitive theories suggest that conceptual understanding is mediated by predominantly visual areas, such as the lateral occipital temporal cortex (LOTC) (Wurm and Lignau, 2015, 2017; Cook and Bird, 2013) that spans regions including the EBA, while motoric theories take the orthogonal view that frontal mirroring regions are involved in coding abstract goal representations in order to mediate action understanding (e.g., Rizzolatti et al., 2014). Hence, the discussion is open-ended as to the organization of the AON and whether this flow of information is truly hierarchical in nature, or more distributed across the network (De Gelder & Solanas, 2021); resultingly, connectivity and organization within these systems as they contribute to identify recognition of the self and others remains unclear.

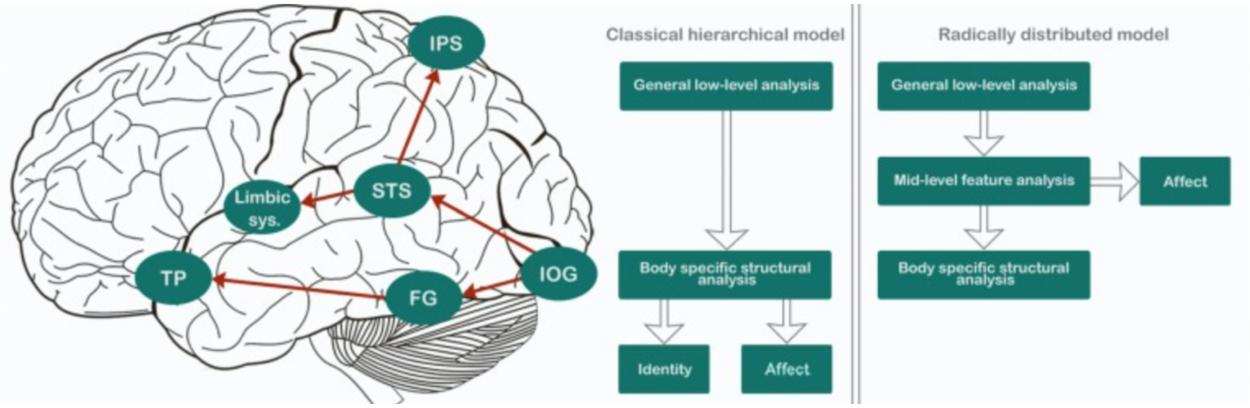


Figure 1.1: Two prevailing models for how conceptual understanding (e.g., identity) is abstracted from bodily featural processing as implemented in the brain. The classic hierarchical model of information processing attributes the function of temporooccipital regions to primarily decode lower-level features of the action that proceeds hierarchically. The radically distributed model suggests that these regions are attuned to mid-level (intermediary) features of the action stimulus that draw insight into higher-level conceptual understanding earlier in processing, which need not proceed hierarchically. *Abbreviations*: FG, fusiform gyrus; IOG, inferior occipital gyrus; IPS, intraparietal sulcus; STS, superior temporal sulcus; TP, temporal pole. Figure and interpretation drawn from De Gelder and Solanas (2021).

1.2.3. Commonalities between action processing of the self and others

The AON, and in particular the mirror neuron system (MNS), reveal how the self and others are linked by a common visuomotor thread, in which perception and action are matched on the same representational platform (Prinz, 1997). The link between actions of the self and the other, is a very difficult computational problem (Barresi & Moore 1996; Saygin and Dick, 2014), rendered simple when reducing the functional mechanism to a core neural substrate. Indeed, from MNS involvement in processing the actions of the self and others, even higher-level social attributions (e.g., personality trait inferences) are grounded upon the system, involving interactions between the MNS and the higher-level, socio-cognitive mentalizing system, consisting of the temporoparietal junction (TPJ), medial precuneus, and the ventromedial (vmPFC) and dorsomedial prefrontal cortex (dmPFC) (Spunt and Lieberman, 2009; 2013; Arioli et al., 2017). This system is traditionally implicated in social cognition tasks spanning a wide range including mental state attribution, perspective taking, and intentionality (Frith and Frith, 2006; Spunt & Lieberman, 2012), and overlaps across motoric and social cognition tasks (Iacoboni et al., 2005). Together, these studies shed light on the multidimensional constructions of both the self and others. The construct of oneself and the construct of "another" are multicomponent, spanning across sensorimotor contingencies and taking on complex, reflexive, and abstract types of cognition, as shown across diverse empirical tasks. Architecturally, the MNS is thought to enact the building block for more abstracted forms of social cognition. It is theorized that higher-level involvement of the higher-level socio-cognitive system, the mentalizing system (Frith & Frith, 2006), is grounded upon the action/perception MNS, and interacts—producing spontaneous dynamic interactions between the two systems required for different social inferences. Indeed, reputed theoretical models (Keysers and Gazzola, 2007) demonstrate a degree of dynamic connectivity between mirroring and mentalizing during social-cognitive tasks that require both simulative motor representations and abstracted, reflective judgments. Importantly, both of these systems together are not just implicated in social cognition, but also for processing of the self, thus indicative of an inherent social and relational construction of the self.

Together, the results across visuomotor to socio-cognitive domains maintain the similar pattern: overlapping neural systems for the self and others (Lombardo et al., 2010). However, it cannot be the case that actions of the self and others are represented identically in the brain. This would lead to an impossible situation— what allows for discrimination and awareness of our own movements from others? Fine-grained differences in person identity, as represented by their geometric distance at more granular levels, are thus required to differentiate between self and others. While representational differences for self and others are often documented within overlapping regions at the socio-cognitive level (e.g., in the medial prefrontal cortex; Courtney and Meyer, 2020), these regions do not typically overlap with those involved in action processing. Hence, at the action level, we still lack representational assessments of the self and others that reveal featural differences between the self and others beyond univariate averaging. Thus, the goal stands in **Chapter 2,** to further measure and connect the representational similarity at the neural level between the actions of the self and others.

1.2.4. Feature sets underlying action processing of self and others

The question as to which features underlie action processing and recognition is largely ill-posed due to the following factors: stimuli differences in empirical work (e.g., using static images versus dynamic motion recognition) (Urgen et al., 2019), the emphasis on either high-level action understanding or low-level action perception as opposed to action recognition (Yovel & O'Toole, 2016), the computational complexity afforded by the large degrees of freedom of the human body (e.g., Pham, Khanh, and Ha, 2014) and many potential view- points and positions of the body that change over time (Thurman and Lu, 2014). Studies investigating the systems-neuroscience of the self typically rely on averaging neural activity across the brain via standard univariate subtractions between task conditions. Evidence from these studies reveal largely overlapping cortical regions for both the self and others (e.g., Lombardo et al., 2010). However, finer-grained methods are required to truly differentiate not just where in the brain these regions are, but the type (i.e., what) information is processed and how these neural patterns reflect identity as represented in these regions. Neural pattern representation typically involves computing featural signatures using encoding metrics, such as representational similarity analysis, which defines which features are critical for self-recognition via measuring embedded correlations between feature-derived similarity matrices and neural activity patterns (Kriegeskorte et al., 2008). Thus, in **Chapter 2**, I will use multivariate encoding metrics (representational similarity analysis) as well as pattern classification decoding methods to ask the following question: what features are common for self and other action processing, and unique to each? Further, across the dissertation (in **Chapters 1** and **3**), I will explore this question not just at the neural level, but also involving behavioral and computational measures related to visuomotor processing.

1.2.4.1 Contributing action features to identity recognition

What features are relevant to action processing of the self and others? In the action processing literature, while a range of critical features contribute to action recognition, they can often divide into two feature sets: form, which captures spatial information such as the posture of an agent, and motion, which captures the temporal dynamics of the agent (either local limb movements or global translational motion) (Giese and Poggio, 2003; Vangeneugden et al., 2014). It is thought that the visuo-cognitive architecture of our brains likely integrates these core signals to ensure the visceral and seamless perception of human actions experienced in daily life. However, active debate remains as to their relative contribution to action processing. It is largely unclear as to whether action processing relies primarily on form (motion derived from form) (e.g., Lange and Lappe, 2006; Lappe 2012) or on motion (form derived from motion) (Mather 2016; Giese & Poggio, 2003). To a large extent, this is due to the neuroanatomical architecture of two distinct visual processing routes involved in biological motion processing: the dorsal (primarily linked to spatial localization and dynamic motion), projecting to the parietal cortex and ventral stream (primarily linked to object and postural information), projecting to the temporal cortex (Ungerleider and Mishkin, 1983). Both form and motion cues are shown to be critical mechanisms involved in identity and action recognition from kinematics (Vangeneugden et al., 2014; Thompson & Parasuraman, 2012). Hence, self-recognition from visually-degraded actions may also rely to a large degree on low and mid-level spatiotemporal features, such as spatial structure of the body, speed, spatiotemporal visual cues to recognize oneself (Coste et al., 2021), or temporal exaggerations of the movement (Hill & Pollick, 2000). Distinctive visual features related to the action are shown to contribute to self-action recognition. Coste et al (2021) recently measured these distinctive features on the basis on movement trajectories ("movement distinctiveness") and identified person-specific kinematic fingerprints that facilitated self-recognition. Similarly other sorts of lower-level features, such as speed of movements or even postural differences may drive the performance advantage for self-recognition from actions, considering their general importance to action recognition and processing of others.

On the other hand, because we have motor experience producing an action, as well as privileged access to motor planning for self-generated movements, it is possible that self-action recognition may be unique, in that it engages specialized motoric mechanisms for the self relative to the processing of others.

1.2.4.2 Contributing action features to social processing

Beyond tightly controlled empirical paradigms, humans reside in complex naturalistic environments that procure rich visual experiences and interactions with others. Hence, empirical paradigms are often encouraged to move beyond the laboratory and capture the dynamics of our natural environment. Importantly, in moving from visually sparse PLDs to naturalistic action stimuli, different sets of neural and featural contributions are likely at play, which can aid in understanding high-level judgments of the visual scene (e.g., Clarke et al., 2005; Lorey et al., 2012). At the neural level, these differences are reflected in much larger recruitment of heteromodal cortical regions as contextual information increases (Haxby et al., 2020; Hasson et al., 2004; Sonkusare et al., 2019), which capture the natural regularities and embeddedness of the self in the environmental milieu (Yeshurun, Nguyen, & Hasson 2021). Mechanistically, however, what could be the featural and visuomotor contributions to the naturalistic processing of actions and social interactions? How does the presence of social context impact these featural representations? These visuomotor features that contribute to naturalistic social interactions are notably different than those involved in isolated actions.

For example, the goal representational space changes as more individuals engage in the interaction and depends on how the interaction partners engage with each other (e.g., co-operatively, competitively, congruently), which thus not only introduces social affordances, but also motor affordances (e.g., Borghi, 2018). Interactions also involve a sense of roleplaying (Goffman 1959) and in fact interaction partners are thought to represent a larger and broadened extension of oneself that together produce relational dynamics acquiring their own autonomy (De Jaegher, Paolo, Gallagher, 2010). Beyond the action "set", the "setting" further interacts and directly influences the interaction. Our social environment is dynamic and replete with visual information, by which the "setting" enacts a top-down role to our social perception, seminally apparent in non-social examples such as visual illusions, where stimulus perception is influenced as a function of context— as well as cross-task across social domains. This leads to a situation of "circular causality" (Bolis & Schillbach, 2019), by which our own actions interact with those of the environment to fit into the environment, or our actions produce effects that change the environment (Levins & Lewontin, 1985; Bolis and Schilbach, 2018; 2019; Friston 2010). Paradigms focused on the importance of the action setting often leverage naturalistic stimuli such as movies to engage multiple stages of perceptual processing (e.g., Haxby et al., 2020), which reveal modulatory effects of context on the processing of different types of action stimuli including social interactions (Kadambi et al., 2020) and agentic action (Haxby et al., 2020). Thus, how do the visuomotor features involved in isolated action processing contribute to action processing across naturalistic and social scenes? At the featural level, what specific role does postural information of the body play, and how does it interact with kinematic information in complex, naturalistic scenes?

Chapter 3 explores this question using advances in computer vision to semantically parse the visual scene and assess the featural contributions across social context.

1.2.5. Intrinsic relations to oneself

Recent empirical advances maintain the importance of avoiding arbitrary task-separation between the self and others in the research design. These studies directly recommend integrating the self and others, via embedding the participant (self) in the research design. This is perhaps most obviously described in second-person neuroscience approaches (Schilbach et al., 2013), where participants (selves) are directly embedded in the research design (e.g., in virtual reality) to naturalistically engage in interactions with others. There is compelling evidence from second-person neuroscience in support of the view of a fundamental difference about the relational component of the interaction, where novel dynamics emerge from the interaction, rather than a separated representational view which is typically ecologically invalid and unfamiliar (e.g., Schillbach et al., 2013; Bolis & Schillbach, 2019). While in the present dissertation, I did not directly embed participants into the interactions, one important aim was to measure participants' intrinsic variability related to social and socio-cognitive difficulties and assess the influence on their recognition of the self and others. These forms of individual variance related to the participant are often treated as noise in empirical paradigms, though numerous studies have revealed how intrinsic variance related to participants often leads to different interparticipant interpretations of the action scene (e.g., Van Den Stock et al., 2015).

In **Chapter 1**, I thus measured individual differences in self-action recognition across a large set of participants (*N* = 101). The following individual differences, linked to both sensorimotor self-processing and others, were measured: motor imagery, schizotypal and autistic traits. Subclinical traits linked to autism and schizotypy were of interest since both conditions are rooted in basic disturbances of the self, as defined in their names: Autism, from the Greek root "autos", Schizophrenia, based on a "disordered self." Both conditions are prevalent in the population and often manifest in social embedding and self-related difficulties in the neurotypical population that affect well-being and societal flourishing.

Aims

In summary, my dissertation aims to reveal the shared and distinct visuomotor mechanisms that underlie processing of actions of the self and others. Bridging behavioral and neurocomputational work, the following questions will serve as guides throughout the thesis: (i) What are the visuomotor mechanisms that underlie action processing of the self and others? (ii) How distinct are neural computations for action recognition of the self and others, and to what degree do they rely on featural processing of form and motion? (iii) How does the presence of naturalistic context impact visuomotor contributions during the perception of social interactions?

References

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. Nature reviews neuroscience, 7(4), 268-277.

Arioli, M., Perani, D., Cappa, S., Proverbio, A. M., Zani, A., Falini, A., & Canessa, N. (2018). Affective and cooperative social interactions modulate effective connectivity within and between the mirror and mentalizing systems. Human Brain Mapping, 39(3), 1412-1427.

Apps, M. A., & Tsakiris, M. (2014). The free-energy self: a predictive coding account of self-recognition. *Neuroscience & Biobehavioral Reviews*, *41*, 85-97.

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. Perception, 33(6), 717-746.

Beardsworth, T., & Buckner, T. (1981). The ability to recognize oneself from a video recording of one's movements without seeing one's body. Bulletin of the Psychonomic Society, 18(1), 19-22.

Blake, R., Turner, L. M., Smoski, M. J., Pozdol, S. L., & Stone, W. L. (2003). Visual recognition of biological motion is impaired in children with autism. Psychological science, 14(2), 151-157.

Blake, R., & Shiffrar, M. (2007). Perception of human motion. Annu. Rev. Psychol., 58, 47-73.

Blanke, O. (2012). Multisensory brain mechanisms of bodily self-consciousness. Nature Reviews Neuroscience, 13(8), 556-571.

Blanke, O., Slater, M., & Serino, A. (2015). Behavioral, neural, and computational principles of bodily self-consciousness. Neuron, 88(1), 145-166.

Bolis, D., & Schilbach, L. (2019). 'Through others we become ourselves': The dialectics of predictive coding and active inference.

Borghi, A. M. (2018). Affordances, context and sociality. Synthese, 1-31.

Calvo-Merino, B., Urgesi, C., Orgs, G., Aglioti, S.M., Haggard, P. Extrastriate body area underlies aesthetic evaluation of body stimuli. Exp. Brain Res., 204 (3) (2010), pp. 447-456

Candidi, M., Stienen, B. M., Aglioti, S. M., & de Gelder, B. (2015). Virtual lesion of right posterior superior temporal sulcus modulates conscious visual perception of fearful expressions in faces and bodies. Cortex, 65, 184-194.

Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. Proceedings of the national Academy of Sciences, 100(9), 5497-5502.

Clarke, T.J., Bradshaw, M.F., Field, D.T., Hampson, S.E., & Rose, D. (2005) The Perception of Emotion from Body Movement in Point-Light Displays of Interpersonal Dialogue. Perception ;34(10):1171-1180. doi:10.1068/p5203

Coste, A., Bardy, B. G., Janaqi, S., Słowiński, P., Tsaneva-Atanasova, K., Goupil, J. L., & Marin, L. (2021). Decoding identity from motion: how motor similarities colour our perception of self and others. Psychological Research. https://doi.org/10.1007/s00426-020-01290-8

Coste, A. Individual Motor Signatures: from human perception to biometric applications. Dimension, 2, 0-5.

Coulson, M. (2004). Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence. Journal of nonverbal behavior, 28(2), 117-139.

Cutting, J. E., & Kozlowski, L. T. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. Bulletin of the psychonomic society, 9(5), 353-356.

Decety, J., Perani, D., Jeannerod, M., Bettinardi, V., Tadary, B., Woods, R., ... & Fazio, F. (1994). Mapping motor representations with positron emission tomography. Nature, 371(6498), 600-602.

de Gelder, B., De Borst, A. W., & Watson, R. (2015). The perception of emotion in body expressions. Wiley Interdisciplinary Reviews: Cognitive Science, 6(2), 149-158.

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

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. Experimental brain research, 91(1), 176-180.

Dittrich, W. H., Troscianko, T., Lea, S. E., & Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. Perception, 25(6), 727-738.

Ferri, S., Rizzolatti, G., & Orban, G. A. (2015). The organization of the posterior parietal cortex devoted to upper limb actions: An fMRI study. Human brain mapping, 36(10), 3845-3866.

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding. Science, 308(5722), 662-667.

Friston, K. (2010). The free-energy principle: a unified brain theory?. Nature reviews neuroscience, 11(2), 127-138.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. Neuron, 50(4), 531-534.

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

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. Trends in cognitive sciences, 2(12), 493-501.

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. Nature Reviews Neuroscience, 4(3), 179-192.

Goffman, E. (1959). The moral career of the mental patient. Psychiatry, 22(2), 123-142.

Goodale M.A., Milner A.D. (1992) Separate visual pathways for perception and action. TINS, 15(1).

Grafton, S. T., & Hamilton, A. F. D. C. (2007). Evidence for a distributed hierarchy of action representation in the brain. Human movement science, 26(4), 590-616.

Grossman ED, Blake R (2002) Brain areas active during visual perception of biological motion. Neuron 35:1167–1175

Grossman ED, Battelli L, Pascual-Leone A (2005) Repetitive TMS over posterior STS disrupts perception of biological motion. Vis Res 45:2847–2853

Gunns, R. E., Johnston, L., & Hudson, S. M. (2002). Victim selection and kinematics: A point-light investigation of vulnerability to attack. Journal of Nonverbal Behavior, 26(3), 129-158.

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

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. science, 303(5664), 1634-1640.

Haxby, J. V., Gobbini, M. I., & Nastase, S. A. (2020). Naturalistic stimuli reveal a dominant role for agentic action in visual representation. Neuroimage, 216, 116561.

Hamilton, A. F., Grafton, S. T., & Hamilton, A. (2007). The motor hierarchy: from kinematics to goals and intentions. Sensorimotor foundations of higher cognition, 22, 381-408.

Heberlein, A. S., Adolphs, R., Tranel, D., & Damasio, H. (2004). Cortical regions for judgments of emotions and personality traits from point-light walkers. Journal of Cognitive Neuroscience, 16(7), 1143-1158.

Hill, H., & Pollick, F. E. (2000). Exaggerating temporal differences enhances recognition of individuals from point light displays. Psychological Science, 11(3), 223-228.

Iacoboni, M., & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. Nature Reviews Neuroscience, 7(12), 942-951.

Iacoboni, M., Koski, L. M., Brass, M., Bekkering, H., Woods, R. P., Dubeau, M. C., ... & Rizzolatti, G. (2001). Reafferent copies of imitated actions in the right superior temporal cortex. Proceedings of the national academy of sciences, 98(24), 13995-13999.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. PLoS biology, 3(3), e79.

Isik, L., Koldewyn, K., Beeler, D., & Kanwisher, N. (2017). Perceiving social interactions in the posterior superior temporal sulcus. Proceedings of the National Academy of Sciences, 114(43), E9145-E9152.

Jastorff, J., Begliomini, C., Fabbri-Destro, M., Rizzolatti, G., & Orban, G. A. (2010). Coding observed motor acts: different organizational principles in the parietal and premotor cortex of humans. Journal of neurophysiology, 104(1), 128-140.

Johnson, K. L., McKay, L. S., & Pollick, F. E. (2011). He throws like a girl (but only when he's sad): Emotion affects sex-decoding of biological motion displays. Cognition, 119(2), 265-280.

Jeannerod, M. (2001). Neural simulation of action: a unifying mechanism for motor cognition. Neuroimage, 14(1), S103-S109.

Jokisch, D., Daum, I., & Troje, N. F. (2006). Self recognition versus recognition of others by biological motion: Viewpoint-dependent effects. Perception, 35(7), 911-920.

Kadambi, A., Ichien, N., Qiu, S., & Lu, H. (2020). Understanding the visual perception of awkward body movements: How interactions go awry. Attention, Perception, & Psychophysics, 82(5), 2544-2557.

Lange, J. (2006). A Model of Biological Motion Perception from Configural Form Cues. Journal of Neuroscience, 26(11), 2894–2906.https://doi.org/10.1523/JNEUROSCI.4915-05.2006

Lappe, M. (2012). Perception of biological motion as motion-from-form. *e-Neuroforum*, *18*(3), 67-73.

Levins, R., & Lewontin, R. (1985). The dialectical biologist. Harvard University Press.

Lorey B., Kaletsch, M., Pilgramm, S., Bischoff, M., Kindermann, S., Sauerbier, I., Stark, R., Zentgraf, K., & Munzert, J. (2012) Confidence in Emotion Perception in Point-Light Displays Varies with the Ability to Perceive Own Emotions. Wenderoth N, ed. PLoS ONE ;7(8):e42169. doi:10.1371/journal.pone.0042169

Loula, F., Prasad, S., Harber, K., & Shiffrar, M. (2005). Recognizing people from their movement. Journal of Experimental Psychology: Human Perception and Performance, 31(1), 210.

Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., Wheelwright, S. J., Sadek, S. A., Suckling, J., ... & Baron-Cohen, S. (2010). Shared neural circuits for mentalizing about the self and others. Journal of cognitive neuroscience, 22(7), 1623-1635.

Masson, H. L., & Isik, L. (2021). Functional selectivity for social interaction perception in the human superior temporal sulcus during natural viewing. NeuroImage, 245, 118741.

Mather, G., Radford, K., & West, S. (1992). Low-level visual processing of biological motion. Proceedings of the Royal Society of London. Series B: Biological Sciences, 249(1325), 149-155.

Parkinson, C., Walker, T. T., Memmi, S., & Wheatley, T. (2017). Emotions are understood from biological motion across remote cultures. Emotion, 17(3), 459.

Pfeifer, J. H., Iacoboni, M., Mazziotta, J. C., & Dapretto, M. (2008). Mirroring others' emotions relates to empathy and interpersonal competence in children. Neuroimage, 39(4), 2076-2085.

Pham, C. H., Le, Q. K., & Le, T. H. (2014). Human action recognition using dynamic time warping and voting algorithm. VNU Journal of Science: Computer Science and Communication Engineering, 30(3).

Pollick, F. E., Lestou, V., Ryu, J., & Cho, S. B. (2002). Estimating the efficiency of recognizing gender and affect from biological motion. Vision research, 42(20), 2345-2355.

Pollick, F. E., Kay, J. W., Heim, K., & Stringer, R. (2005). Gender recognition from point-light walkers. Journal of Experimental Psychology: Human Perception and Performance, 31(6), 1247.

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

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. Nature reviews neuroscience, 11(4), 264-274.

Roether, Claire L., et al. "Critical features for the perception of emotion from gait." Journal of vision 9.6 (2009): 15-15.

Saygin, A. P. (2007). Superior temporal and premotor brain areas necessary for biological motion perception. Brain: a Journal of Neurology, 130(9), 2452e2461.

Saygin, A. P. (2012). Biological motion perception and the brain: Neuropsychological and neuroimaging studies. In K. Johnson, & M. Shiffrar (Eds.), People watching: Social, perceptual, and neurophysiological Studies of body perception. Oxford series in visual cognition. Oxford University Press.

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

Sonkusare, S., Breakspear, M., & Guo, C. (2019). Naturalistic stimuli in neuroscience: critically acclaimed. Trends in cognitive sciences, 23(8), 699-714.

Spunt, R. P., & Lieberman, M. D. (2012). Dissociating modality-specific and supramodal neural systems for action understanding. Journal of Neuroscience, 32(10), 3575-3583.

Thompson, J., & Parasuraman, R. (2012). Attention, biological motion, and action recognition. Neuroimage, 59(1), 4-13.

Thoresen, J. C., Vuong, Q. C., & Atkinson, A. P. (2012). First impressions: Gait cues drive reliable trait judgements. Cognition, 124(3), 261-271.

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

McAleer, P., Pollick, F. E., Love, S. A., Crabbe, F., & Zacks, J. M. (2014). The role of kinematics in cortical regions for continuous human motion perception. Cognitive, Affective, & Behavioral Neuroscience, 14(1), 307-318.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. Trends in neurosciences, 6, 414-417.

Van den Stock J, Hortensius R, Sinke C, Goebel R, de Gelder B. 2015. Personality traits predict 1345 brain activation and connectivity when witnessing a violent conflict. Scientific Reports 5. 1346 doi:10.1038/srep13779

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

Verosky, S. C., & Todorov, A. (2010). Differential neural responses to faces physically similar to the self as a function of their valence. NeuroImage, 49(2), 1690-1698.

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

Yeshurun, Y., Nguyen, M., & Hasson, U. (2021). The default mode network: where the idiosyncratic self meets the shared social world. Nature Reviews Neuroscience, 22(3), 181-192.

Yovel, G., & O'Toole, A. J. (2016). Recognizing people in motion. Trends in cognitive sciences, 20(5), 383-395.

Zimmermann, M., Mars, R. B., De Lange, F. P., Toni, I., & Verhagen, L. (2018). Is the extrastriate body area part of the dorsal visuomotor stream?. Brain Structure and Function, 223(1), 31-46.

CHAPTER 2

Motoric and intrinsic factors modulate self-recognition of actions

Abstract

Although humans can recognize their body movements in point-light displays, self-recognition performance varies substantially across action types and participants. Are these variations primarily due to an awareness of distinct movement patterns, or to underlying factors related to motoric planning and/or participant traits? To address these questions, we conducted a large-scale study in self-action recognition ($N = 101$) and motion captured whole-body movements of participants who performed 27 different actions across action goals and degree of motor planning. After a long delay period (~ 1 month), participants were tested in a self-recognition task: identifying their point-light action amongst three other point-light actors performing identical actions. While we found no relationship between self-recognition performance and movement distinctiveness, we found that performance was modulated by interactions between extrinsic factors (associated with the degree of motor planning in performed actions) and intrinsic traits linked to individuals' motor imagery ability and subclinical sensorimotor self-processing (autism and schizotypal traits). These interactions show that recognition of our own actions is influenced by factors beyond movement idiosyncrasies, with mechanistic implications for *how* the motor system may augment vision to construct self-awareness.

Introduction

Recognition of oneself depends on more than visual experience. As a prime example, humans can recognize their own actions even from minimal visual input provided by point-light action displays (Cutting & Kozlowski, 1977; Loula et al., 2005; Beardsworth & Buckner, 1981)—disconnected dot animations depicting the motion of key joints on the human body (Johansson, 1973). Despite removing body identity information and depicting actions in an unfamiliar third-person perspective, self-recognition from point-light movements suggests that people can readily connect motor experience of performing actions to visual recognition of actions. These findings are consistent with multiple lines of evidence that have revealed close links between motor and visual representations of actions (e.g., Binet, 1886; Viviani & Stucchi, 1992; Prinz, 1997; Knoblich & Flach, 2001; Casile & Giese, 2006; Rizzolatti, Fogassi, & Gallese, 2001; Blake & Shiffrar, 2007; Iacoboni 2008).

Previous research has identified a few core characteristics of self-recognition from actions. From point-light actions, participants are more accurate in recognizing themselves than visually familiar friends (Cutting & Kozlowski, 1977), even when configural topographic cues are disrupted (Cook, Johnston, & Heyes, 2012; cf. Loula et al., 2005), or visual feedback is removed (Casile & Giese, 2006; Bläsing and Sauzet, 2018). People can identify their own actions from whole-body movements (Loula et al., 2005) and even from hand movements (Conson, Aromino & Trojano, 2010). Self-recognition of one's own gait is viewpoint-invariant (cf. Mitchell & Curry, 2016), but recognition of familiar friends from their gait patterns is viewpoint-dependent, with better performance from frontal than profile views (Jokisch, Daum & Troje, 2006). View-independent performance for self-recognition from point-lights has also been observed for actions less common than walking (Loula et al., 2005; Burling et al., 2019), suggesting that the visual representation of self-actions based primarily on motor experience is fundamentally different from action representations of others learned from visual experience. The self-action advantage also extends beyond explicit visual recognition of whole-bodies, influencing the predictive accuracy of future action outcomes (e.g., Knoblich & Flach, 2001; Bischoff et al., 2012; cf. Hohmann et al., 2011), task-irrelevant (implicit), body part recognition (Frassinetti et al., 2009; Daprati & Sirigu, 2002; Daprati, Weriessneggeer, & Lacquaniti, 2007), facial expression recognition (Cook, Johnson & Heyes, 2012), memory for self-performed action verbs (i.e., enactment) (Engelkamp & Krumnacker, 1980), and multimodal action recognition (e.g., Repp & Knoblich, 2004; Flach, Knoblich, & Prinz, 2004; Murgia et al., 2012; Kennel et al., 2014).

While these studies reveal that self-action recognition can arise from motor experience despite lacking visual experience, the underlying mechanisms are yet to be clarified. One empirical approach has been to include a range of actions in the paradigm, as self-recognition performance is known to vary substantially across different actions. Loula et al. (2005) recorded ten actions from each individual participant and asked participants to report the identity (self, friend or stranger) of the point-light actors. The researchers found that identity performance varied significantly across actions in the range of 40~80% (chance level of 33.3%). For example, people were more accurate in self-recognition for complex actions such as dancing (~80%) and boxing (~65%), than for relatively simple and routine actions such as walking/running (~40%). Similarly, Sevdalis and Keller (2009) found greater self-recognition performance for dancing movements than for stereotyped gait patterns. In a recent study by Coste et al (2021), the variability in self-recognition performance was measured and attributed to movement similarity conveyed by their own body motions versus actions performed by others. In this study, participants were asked to perform postural motions with general instructions such as "*create postural motions by keeping* *knees extended with toes and heels in constant contact with the floor*." Although people recognized their own actions above chance, people often misattributed movements of others to themselves that were highly visually similar. These results are indicative of a "kinematic fingerprint", in that people recognize themselves based on the degree of visual similarity to their own kinematic styles of body movements, as a type of familiar idiosyncratic movement feature for each person. However, the study gave participants the same general instructions for body movements in performed actions without varying the action goals or action complexity. Beyond visually similar properties of our own actions and actions performed by others, the influence of other potential contributors to self-recognition accuracy is unclear.

Notably, other factors have also been shown to be especially important in influencing action processing of the self and others. In the action-processing literature, the motoric goal complexity of the action modulates the degree of motor planning required for action production (e.g., Burling et al., 2018) (complex goals require more motor planning than simple goals) and evokes different neural activity in the action observation network even when the same effectors are used (Molnar-Szakacs et al., 2006; Li et al., 2020). Although most previous studies in self-action recognition opt for group-level averaging, variability in individual differences has also been shown to influence self-action recognition (e.g., Burling et al., 2018). At the individual level, self-processing is characterized as atypical in clinical conditions linked to the sensorimotor system, including the Autism Spectrum Conditions (ASC) – from the Greek root "autos" for self (Kanner 1943; Asperger, 1944; Perrykkad & Howhy, 2020), and Schizophrenia Spectrum Conditions (SSC), a "disordered self", characterized by "a disunity of consciousness" (Bleuler, 1911; Schneider 1959; Sass & Parnas, 2003; Kraeplin 1919; Parnas & Herniksen, 2014; Sandsten et al., 2020). In these conditions, atypical self-representations are present and primarily attributed to an early disturbance at the bodily level (e.g., Noel et al., 2016; Sass & Borda, 2015; Gallese & Ferri, 2014; Gessaroli et al., 2013; Perrykkad and Howe, 2020). These differences manifest in opposite ways in ASC and SSC (Noel, Cascio, Wallace, & Park, 2016), related to either sharpened or blurred internal self/other sensorimotor boundaries, respectively (Noel et al., 2016; Fletcher & Frith, 2009). Importantly, the differences extend to the general population. Participant-level variability linked to these conditions affects self-identification (Mul et al., 2019) and spatial self-boundaries (Ferroni et al., 2020), and further impacts the general processing of action stimuli (Miller & Saygin, 2013; Puglia & Morris, 2017; Thurman et al., 2016; van Boxtel et al., 2017; Wang et. al., 2018; Chakraborty & Chakrabarti, 2015) and even in self-action recognition from point-light displays (autistic traits: Burling et al., 2018). Hence, while one potential form of variability in self-recognition accuracy could indeed result from distinctive kinematic styles of own body movements as in Coste et al., 2021, another—perhaps deeper source of variance, may be driven by differences in motor planning required by different action goals, and/or individual differences in self-processing.

To assess the influence of intrinsic participant and extrinsic motoric factors that may extend beyond visual distinctiveness, we conducted the first large-scale study (*N* = 101) in self-action recognition and asked the following: is self-recognition performance primarily reflected in idiosyncratic motion of individuals? Or does self-recognition depend on systematic properties from extrinsic factors of the action and/or intrinsic traits of the participant? To measure the influence of extrinsic factors, we included a range of actions, varying action complexity based on the degree of motor planning involved in performing actions. To measure how movement distinctiveness across individuals influences self-recognition performance, we implemented a widely-used algorithm in spatiotemporal signal processing— dynamic time warping (DTW)

(Gavrila & Davis, 1995)— to quantify action similarity between self-movements and actions performed by others based on movement trajectories of joints in actions. To measure the influence of participant-level intrinsic traits, we computed the relation between self-recognition performance and three individual difference measures linked to sensorimotor self-processing: autistic traits, schizotypal traits, and motor imagery ability. Motor imagery was included as a third intrinsic measure since motor imagery serves as a part of the action simulation network, triggered either through visual observation of others or through self-imaging. Moreover, action recognition and motor simulation are well-characterized in the literature as sharing overlapping neural and behavioral resources (Jeannerod & Pacherie, 2004) and self-processing is shown to recruit motor imagery mechanisms (e.g., Conson et al., 2010).

The experimental design consisted of two sessions. In the first session, each participant performed 27 actions that were recorded through motion capture. Participants were informed that their actions were recorded for an action recognition study but were importantly never informed of the study's focus on self-recognition. After a delay of about one month on average, participants returned to complete a self-recognition task that required identifying their own action among three other actors performing the same action. We introduced three types of actions during the motion capture session: nine simple actions and nine complex actions that were verbally instructed (i.e., "please naturalistically perform the action: *to grab / to get attention*"), and nine actions that were provided with video instruction, asking participants to imitate body movements of an actor shown in a video. We included video instruction since imitation elicits a unique action requirement: action imitation reduces motor planning demands, due to copying another's motor sequence (see enactment effect; Zimmer & Engelkamp, 1996). If self-recognition performance depends primarily on the degree of motor planning involved in performing actions, we would expect that imitation actions may yield lower accuracy in identifying own body movements than complex actions provided with verbal labels, even at comparable levels of action complexity and across large variability in performing the same actions across individuals. If, however, distinctiveness in performing actions is the primary contributor to self-recognition performance, any performance differences between the action types should be attributed to visual properties of the stimulus, such as movement distinctiveness or speed.

In summary, we hypothesized that participants would not only recognize their own actions in point-light displays, but that recognition performance would vary systematically, according to underlying extrinsic factors – action types linked to differences in motor planning– and intrinsic factors of participants, linked to sensorimotor self-processing.

Method

Participants. 108 undergraduate students ($M_{\text{age}} = 21.20$, $SD_{\text{age}} = 3.81$, females = 79, males = 29) were recruited through the Subject Pool at the University of California, Los Angeles. Sample size was determined sufficiently sensitive in accord with a previous study on the relation between emotion discrimination from gait patterns and autistic and schizotypy traits (Blain, Peterman, and Park, 2017). The study was approved by the UCLA Institutional Review board. All participants were provided course credit for their participation and were naïve to the purpose of the study. Participants had normal or corrected-to-normal vision and no physical disabilities. Seven participants were excluded due to inputting errors of participant motion capture files, resulting in a total of 101 participants included in the analysis (M = 28, F = 73).

Apparatus. Participants' body movements were recorded using the Microsoft Kinect V2.0 and Kinect SDK in a quiet testing room. Participants were instructed to perform the actions in a rectangular 0.76 m x 1.52 m space, in order to provide flexibility to perform the action, while remaining within recording distance. The Kinect was placed 1.52 m above the floor and 2.59 m away from the participant. The three-dimensional (X-Y-Z) coordinates of the key joints were extracted at a rate of approximately 33 frames per second and later used to generate point-light displays of actions (see Figure 2). Customized software developed in our lab was utilized to enhance movement signals, and to carry out additional processing and trimming for actions presented later in the testing phase (van Boxtel & Lu, 2013).

Stimuli Generation. All actions were selected to be commonly encountered actions and were captured by the motion capture system. The actions varied in complexity in order to characterize a broad range of common movements in daily life. First, participants were instructed to freely perform nine simple and nine complex actions provided with clear verbal labels (i.e., verbal instruction). Participants indicated the start/stop of their action with a T-position. Simple and complex actions were extended from Burling et al (2019), delineated by whether the action was a simple goal (e.g., *wave*), conveying a routine action with less motor planning, or a complex goal (e.g., *argue*) linked to more detailed motor planning. The simple actions included *grab, jump, wave, lift, kick, hammer, push, point, punch.* The complex actions included: *argue, macarena, wash windows, play baseball, get attention, hurry up, fight, stretch,* and *play guitar*.

In addition, participants also performed actions based on video instruction. Participants were instructed to view an action performed by actors shown in a stick-figure display (i.e., lines connected between joint positions) on a computer, and subsequently imitate the seen action afterwards. Nine imitation actions were selected from the Carnegie Mellon Graphics (CMU) Lab Motion Capture Database available online (http://mocap.cs.cmu.edu), generated from pre-defined actors. The actions were selected to capture a broad range of goal variability: *jumping jacks,* *basketball, bend, direct traffic 1, direct traffic 2, conversation, laugh, digging a hole, and chopping wood.* To account for any implicit goal-attribution or familiarity with action sequences that could impact the degree and/or type of imitation (e.g., Bekkering et al., 2000; Wöhlschalger et al., 2003; Wild et al., 2009), we also included a range of familiar (e.g., bend) and unfamiliar (e.g., directing traffic) imitation action sequences. Participants were never provided the verbal label for imitation actions. Each video displayed an actor shown as a stick figure performing one of the imitation actions and was presented in three different angles to the subject, either to the right or left (+/- 45°; half-profile) or facing forward (0°; frontal) by rotating the horizontal axis. Each imitation action was recorded twice: once after viewing the three different angles, and once more after viewing only the forward-facing angle. The first imitation recording was discarded (served as practice during motion recording), and only the second imitation recording was used in the self-recognition experiment. The recorded raw motion data from the Kinect system were passed through a double exponential adaptive smoothing filter (LaViola, 2003) to remove noisy and jittered movements (e.g., ballistic random jumps of points). Additionally, the stimuli were trimmed and processed to display the point light-displays using BioMotion Toolbox (van Boxtel & Lu, 2013) with their segmented action recording, which would be reiteratively looped in the self-recognition session.

Procedure

The experiment was split into two phases: motion recording and recognition testing. The first phase consisted of a motion recording session, where participants performed the actions and were recorded with a motion capture system. Participants were informed that their actions were recorded for an action recognition study, but were never informed about the aim of the study. After a delay period (M = 37.39 days, SD = 5.20 days; range: 23 – 56 days), participants returned for the second phase to complete two recognition tasks. In the first task, participants completed a self-recognition task by identifying their own actions that were recorded with the motion capture system. In the second task, participants completed a visual recognition task, where they identified the actions imitated during the motion recording session (further detailed below). The order of the self-recognition task and the visual recognition task was counterbalanced between participants.

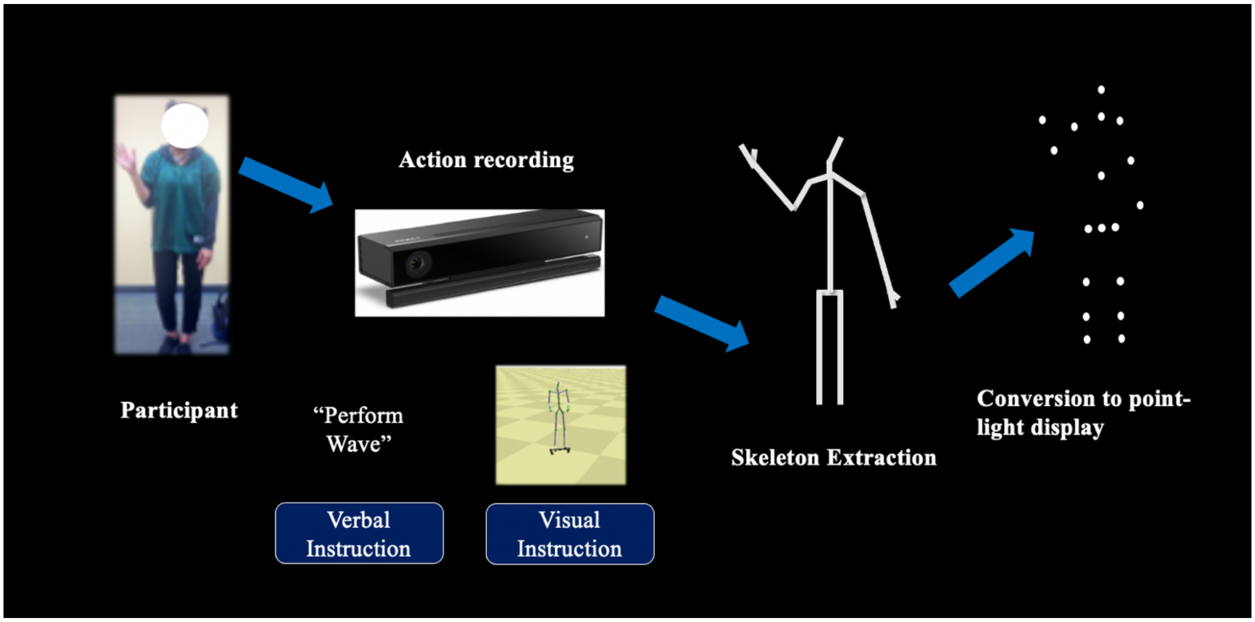


Figure 2. Illustration of motion recording (session 1). Actions were recorded by the Kinect system and converted to point-light displays. Participants were instructed to perform the actions either through verbal instruction (simple, complex) or visual instruction guided by a stick figure (imitation).

Motion Recording Phase

In the first phase (recording session), participants were provided verbal instruction for the 18 actions (nine simple and nine complex) and asked to perform the actions as naturally as possible (i.e., "please naturalistically perform the action: *to grab*"). For the remaining nine imitation actions, none of the participants were provided the verbal label of the action. Instead, participants were instructed to *imitate* the movements of the action presented in a stick figure video (*Figure 2*).

Upon completing the action recording, participants completed two questionnaires: Schizotypal Personality Questionnaire (SPQ; Raine, 1991) and the revised Vividness of Motor Imagery Questionnaire (VMIQ-2; Roberts et al., 2008). The SPQ was administered to assess degrees of schizotypal traits among individuals in the typical population. The VMIQ-2 was included to assess motor imagery differences as a potential source of variability in self-processing and biological motion processing.

Recognition Phase: Self-Recognition Task and Visual-Recognition Task

In the second phase (recognition test), participants were seated approximately 0.76 meters in front of a monitor in a dimly lit room and were asked to select their own action amongst three other distractor actions spread out horizontally along the center of the screen, as shown in Figure 3. Each action was presented with 17 point-lights located at key joints, in three different orientations (rotated around the vertical axis) 0°, (frontal), 45° (half-profile, right), 225° (half-profile, left), for a total of 81 trials. All of the actions within a trial displayed the same orientation. Participants were instructed to select their own point-light action amongst four displays. The actions were looped until the participant selected one of the point-light actors, each depicted in one of four horizontally-spread, randomly arranged boxes, or until a time limit of 30 seconds. The four animations included their own action and the same action performed by three distractor actors with the same gender, all of whom participants were unfamiliar with. All the point-light actions were normalized for maximum height. Following selection of the action, participants were asked to provide a confidence judgment, in which they rated the confidence of their selection from 1 (not at all confident) to 5 (most confident). Participants were not provided any feedback. Following the recognition task, participants were asked to complete the Autism-Spectrum Quotient (AQ) questionnaire to assess their degree of Autistic traits (Baron-Cohen et al., 2001).

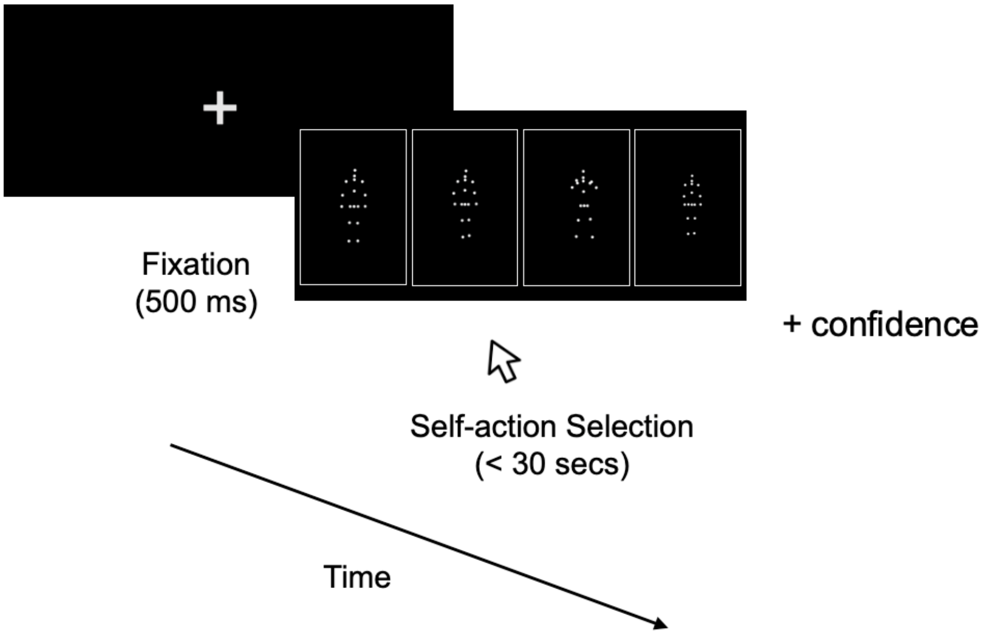


Figure 3. Schematic illustration of a sample trial showing "push" action for the self-recognition task. One point-light action is the participant's action, while the other three point-light actions are distractors performed by gender-matched actors. During the display of actions, participants were provided a maximum of 30 seconds to click on the action that was their own. Following the actions, participants rated their overall confidence in the selection (1: not at all confident to 5: most confident)

Individual Difference Measures

Autistic Quotient We assessed self-reported autism-spectrum traits in the general population using the Autism-Spectrum Quotient (AQ) questionnaire, consisting of 50 questions designed to measure five different subtypes: *social skill, attention switching, attention to detail, social communication,* and *imagination* (Baron-Cohen et al., 2001). Response criteria requires the selection of one of the four possibilities (four-point scale): "definitely disagree", "slightly disagree", "slightly agree", "definitely agree." While not a diagnostic instrument, scores of 32+ on the AQ in the general population are generally indicative of a predisposition to ASC, out of a maximum score of 50 points (1 point per question validating autism-spectrum traits), though the cutoff for ASC individuals on the AQ is typically shown to be greater than 26 (Woodbury-Smith, Robinson, Wheelwright, & Baron-Cohen, 2005; Ruzich et al., 2016).

Schizotypal Personality Questionnaire To measure trait-variance related to Schizotypal Personality Disorder amongst neurotypics, we used the Schizotypal Personality Questionnaire (SPQ) developed by Raine (1991). The 74-item survey was based on criteria from the *DSM-III-R* (American Psychiatric Association, 1987) that measures schizotypy from multiple dimensions (positive, negative, disorganized, and paranoia) and captures its phenotype, etiology, symptomatology (Wuthrich and Bates, 2006; Barron et al., 2015). We chose the full-scale SPQ (74 items) rather than the more recent SPQ-brief (32 items; Cohen et al., 2010), as the original SPQ has been shown to provide a clearer division of the individual subscales (Ford and Crewther, 2014), important when considering the reflected overlap between AQ and SPQ subscales. The SPQ adopts a three-factor structure (analogous to the symptom structure in Schizotypal Personality Disorder and SSC), measuring three main constructs of schizotypy: the cognitive-perceptual dimension (positive schizotypy), interpersonal dimension (negative schizotypy), and disorganized feature dimension (disorganized schizotypy) based on *DSM-III-R* criteria (Raine, 1994), but well-matched to current *DSM-V* criteria (Barron et al., 2015). Further divisions within the three-factor structure include nine different subscales of the SPQ: *ideas of reference, unusual perceptual experiences, odd/magical beliefs, suspiciousness/paranoid ideation* (cognitive-perceptual)*; social anxiety, no close friends, constricted affect* (interpersonal); *odd behavior and appearance,* and *odd speech* (disorganized) (Raine 1991; 1994). Response criteria on the questionnaire requires binary selection of “true” or “false” (two-point scale) to particular statements (e.g., “I am aware that people notice me when I go out for a meal or to see a film”). The top 10% of scorers typically reflect scores greater than 41, while the bottom 10% typically score 12 or lower.

Vividness of Motor Imagery Questionnaire The VMIQ-2 (Roberts, 2008) measures introspective reports of vividness of imagery in kinesthetic (movement simulation), internal (first person simulation), and external (third person simulation) visual imagery of 12 different actions (e.g., kicking). Vividness of motor imagery is rated on a five-point Likert scale for each of the 12 actions in each of the three sub-areas. Note that lower scores in VMIQ-2 indicate more vivid images and stronger motor imagery ability.

Action similarity measures

People perform some actions with highly similar movements across individuals, but other actions with distinctively different body movements. To assess the role of movement distinctiveness as a contributor to self-recognition, we implemented the dynamic time warping (DTW) algorithm. DTW allows for quantitative measures of similarity between two temporal sequences based on motion trajectories in actions. The algorithm uses a non-linear, monotonic temporal warping algorithm to measure the similarity between time series data such as movement trajectories of joints (Gavrila & Davis, 1995). DTW is also able to account for the time-variance across human actions, in how different people may take different amounts of time with different movement trajectories even when performing the same action. The DTW algorithm is designed to find an optimal match (warping path) for temporal alignment between these pairwise time sequences that minimizes their dissimilarity regardless of low-level factors (e.g., speed, duration differences). Smaller DTW distance values indicate greater similarity of body movements between participants when they performed the same action. A pseudo-code of DTW algorithm can be found in (Pham, Le, & Le, 2014).

The following steps were implemented for DTW analyses:

(1) For each participant's action, we extracted the 3D positions of each of the 17 joints using the BioMotion toolbox (van Boxtel & Lu, 2013).

(2) We centered each trajectory of a joint to zero in order to remove the impact of global factors (e.g., global body displacements, limb length, etc.) on the similarity measures.

(3) We then implemented an action DTW algorithm (Pham, Le, & Le, 2014) to search for a temporal warping function shared across all 17 joints. See Figure 4 for the illustration of temporal warping for joint trajectories performed by two actors in the action DTW algorithm.

(4) After deriving the optimal warping function, the analysis computes the frame-by-frame Euclidean distances of the temporally warped joint trajectories in actions performed by different actors.

(5) DTW distance was then computed as the sum of the distances between all joint trajectories normalized by the number of frames of a target actor. This normalization step is required in order to account for the different durations across participants performing the same action

(6) For each participant, the dissimilarity of the target participant performing an action from all other participants was captured by a mean DTW distance measure, computed by averaging across pairwise DTW distances between the target participant with all other participants in performing this action.

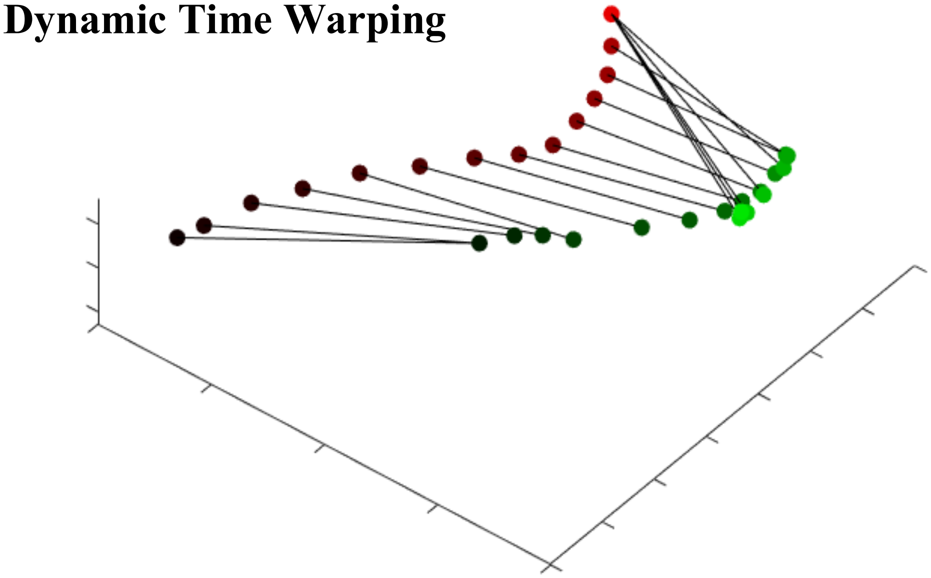


Figure 4. Illustration of DTW alignment via temporal warping for joint movements (sample joint: elbow) in an action (sample action: wave) between performed by two actors (red or green). The gradient in circle colors indicates the elapse of time, with darker colors reflective of earlier frames, and brighter color in later frames. The temporal warping function is illustrated as connected lines.

Results

Impact of extrinsic factors (action types) on self-recognition performance

As shown in Figure 5, participants were able to recognize self-actions significantly above chance performance (0.25 in a 4AFC task): for simple actions with verbal instruction (*M* = 0.41, *SD* = 0.16), *t*(100) = 9.45, *p* < .001, cohen's *d* = .940, for complex actions with verbal instruction (*M* = 0.54, *SD* = 0.17), *t*(100) = 16.86, *p* < .001, *d* = 1.678, and for imitated actions without verbal instruction (*M* = 0.39, *SD* = 0.16), *t*(100) = 9.20, *p* < .001, *d* = .916, corroborating prior research (e.g., Burling et al., 2019; Loula et al., 2005). Confidence data further showed consistent results with self-recognition accuracy for the action types (see Appendix for further detail).

A two-way repeated-measures ANOVA with action orientations and action types as within-subjects factors on self-recognition performance revealed no effect of action orientation, facing left: 225° (*M* = 0.45, *SD* = 0.16), front: 0° (*M* = 0.45, *SD* = 0.14), right: 45° (*M* = 0.44, *SD* = 0.15), *F*(2, 200) = 1.58, *p* = .207, η_p^2 = .016. The finding is consistent with a previous study showing that self-recognition of walking actions is independent of the viewing angle (Jokisch, Daum, & Troje, 2004). However, the analysis yielded a significant main effect of action type (simple, complex, and imitation), *F*(2, 200) = 43.09, $p < .001$, η_p^2 = .301. Bonferroni-corrected pairwise comparisons further revealed that self-recognition was more accurate for complex than simple actions (*t*(100) = 7.86, $p < .001$, *d* = .782) and imitation actions (*t*(100) = 8.56, $p < .001$, *d* = .851). No difference emerged between simple and imitation actions (*t*(100) = 0.54, *p* = .593, *d* = .053). We also found a significant two-way interaction between orientation and action type (Figure 5), *F*(4, 400) = 2.77, *p* = .027, η_p^2 = .027, mainly driven by the weaker performance in the frontal view than side views for the imitation actions (*F*(2, 200) = 4.124, *p* = .018, η_p^2 = .040). However, the small effect size calls for caution in interpretation of this interaction effect, as the degree of visual ambiguity due to body occlusion likely impacted self-recognition performance for certain action orientations. Among the imitation actions, three actions (e.g., bend, digging hole and chopping wood) consisted of the most self-occlusion from the frontal view, resulting in grouping ambiguity in the point-light display.

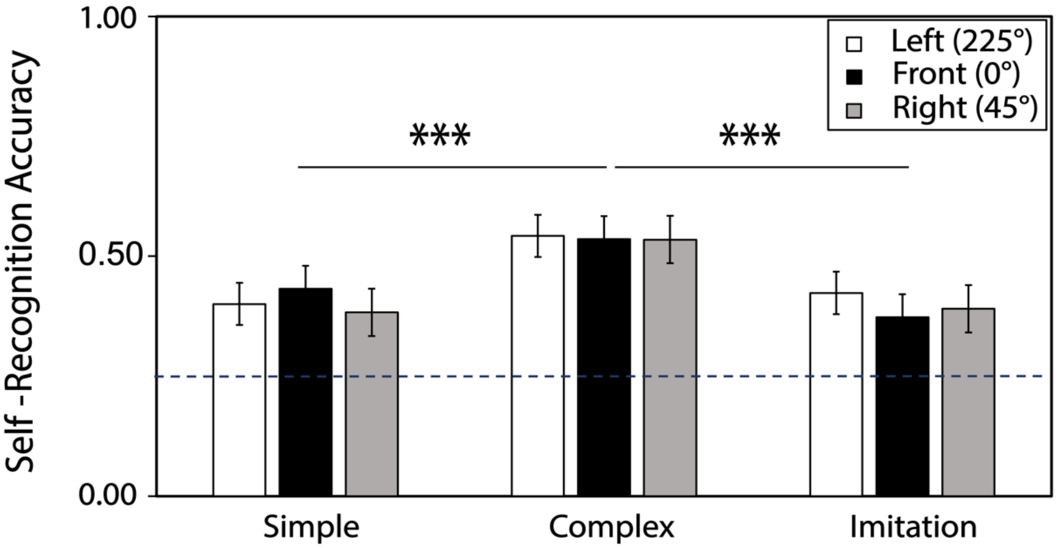


Figure 5. Results of self-recognition accuracy by action types (x-axis) and orientation (colored bars). All action types recognized significantly above chance regardless of viewpoint. Complex actions showed greater self-recognition performance than did simple or imitation actions. Dashed line indicates chance performance (0.25). The error bars indicate standard error of means.

Impact of individual actions on self-recognition performance

Analysis of item-level variability in self-recognition performance for individual actions (Figure 6) revealed that self-recognition performance varied in a large range of .27 to .59 across the 27 actions (with the chance level of .25 for recognizing self from four actions in the experiment). All complex actions (*stretch*, *get attention, wash windows, argue, guitar, hurry up, fight, baseball,* and *macarena*) were self-recognized significantly above chance (*ps* < .001). Most simple actions (*point, punch, lift, grab, push, hammer, jump,* and *kick*) except *wave*, and most imitation actions (*basketball, conversation, directing traffic p1, directing traffic p2, chopping, digging,* and *laugh*) except *bend*, were also self-recognized significantly above chance performance (*ps* < .030). Due to variability in performance at the individual action level, we ran an analysis with a linear mixed effects model, including "individual actions" and "participants" as random effects and "action type" as a fixed effect. The model accounted for residual variance in the individual actions across participants, with a significant main effect of action type on self-recognition performance (*F*(2.61, 1.31) = 11.91, *p* < .001), confirming the impact of the extrinsic factor based on action type on self-action recognition.

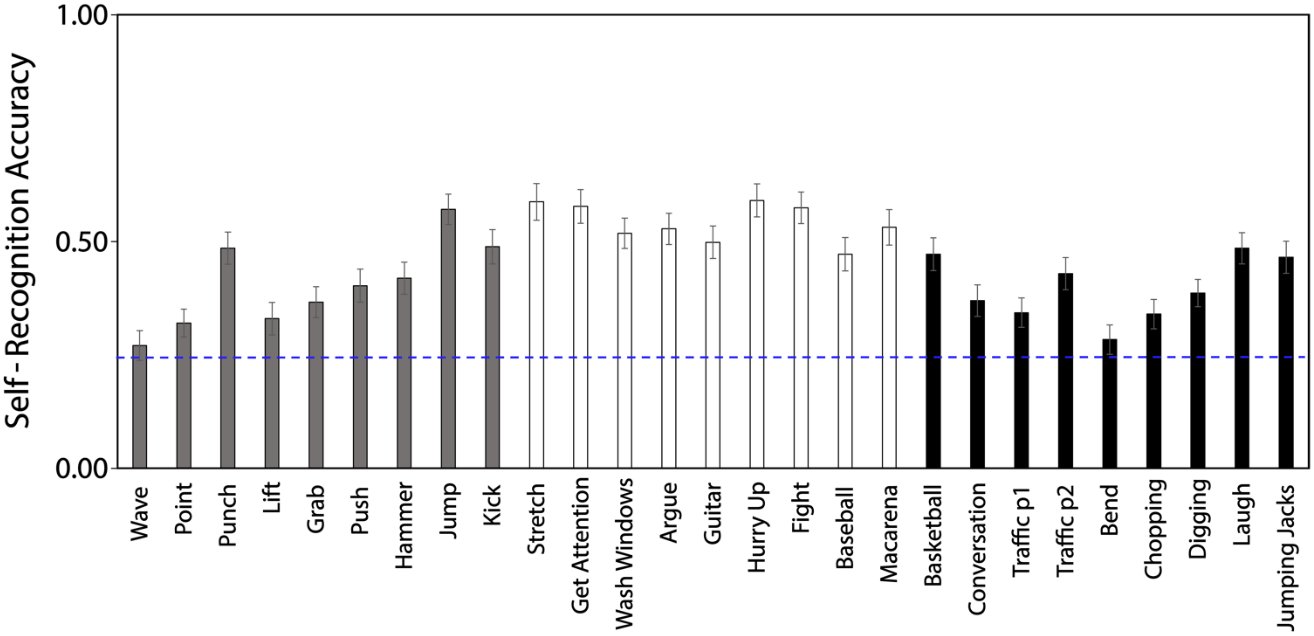


Figure 6. Results of self-recognition accuracy as a function of the individual action. Color-coding of bars corresponds to action type: simple actions denoted in gray, complex actions denoted in white, imitation actions denoted in black. Dashed line indicates chance performance (0.25). Error bars denote standard error of the means.

No influence of movement distinctiveness on self-recognition performance

In addition to the influence of the action type on self-recognition performance, we then measured whether the performance difference was driven by the movement distinctiveness, measured as the dissimilarities between how people perform the same actions. To do so, we examined the impact of movement dissimilarity on self-recognition accuracy using a well-known dissimilarity measure to quantify motion trajectories in actions, dynamic time warping (DTW). For each action, we computed the DTW distance between the action performed by the participant and the movements of other participants performing the same action. Greater DTW distance indicates higher dissimilarity of movements. First, movement dissimilarity showed clear differences in terms of action DTW distances across different action type, as shown in Figure 6. Specifically, the repeated-measures ANOVA revealed a significant main effect of action type on DTW distances, *F*(2,200) = 340, $p < .001$, $\eta_p^2 = .773$. Paired-sample T-tests further revealed a significant decrease in DTW distances for simple actions (*M* = 1.55, *SD* = 0.21) than for complex actions (*M* = 1.97, *SD* = 0.23), *t*(100) = 20.840, $p < .001$, *d* = 2.07, showing that movement trajectories in simple actions were significantly more similar across participants than complex actions. Movements in simple actions were also confirmed as more similar to each other than imitation actions (*M* = 1.99, *SD* = 0.17), *t*(100) = 26.857, $p < .001$, *d* = 2.67. Note that action DTW distances were not significantly different between complex and imitation actions in contrast to the difference in self-recognition performance, in which complex actions were self-recognized at greater rates than imitation actions. Similarly, in contrast to movement dissimilarity, the self-recognition performance pattern was similar for simple and imitation actions, but significantly greater for complex actions. Hence, movement dissimilarity does not provide a satisfying account for the pattern of self-recognition performance for the action types.

The null relationship between self-recognition performance and movement dissimilarity was further confirmed using linear mixed modeling. First, we computed the correlation between self-recognition performance and movement dissimilarity measured as action DTW distances across participants and action types. The correlation was not significant (spearman $\rho = .078$, $p = .174$). DTW dissimilarity for each action type was then entered as a random effect to the linear mixed model previously run on the action types. The full model thus included “individual actions”, “participant”, and “DTW” (dissimilarity value for each action type) as random effects, and action type (simple, complex, imitation) as a fixed factor. The contribution of DTW was tested using Log-Likelihood by comparing the fit of the full model including DTW and the reduced model without DTW. No effect of DTW on self-recognition accuracy was found, $\chi^2(1) = 2.635$, $p = .104$, though the model showed significant influence of the other two random factors on self-recognition performance (individual action: $\chi^2(1) = 49.66$, p < .001; participant: $\chi^2(1) = 64.40$, $p < .001$). Together, the pattern of results rules out the possibility that self-recognition performance is primarily determined by the dissimilarity of body movements of individuals who perform actions in different ways from others.

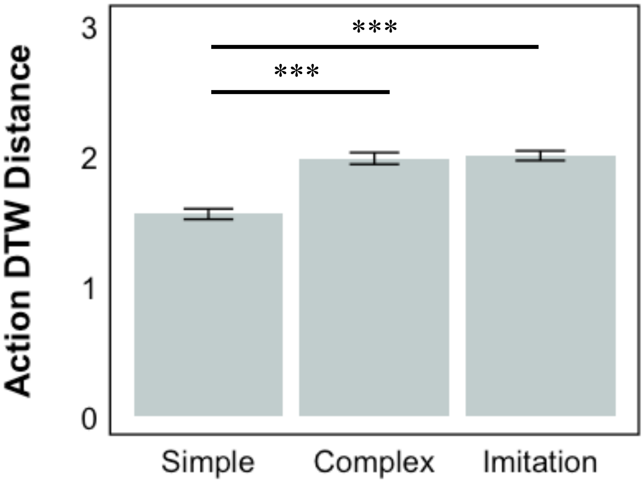


Figure 7. Action DTW distance for different action type, showing simple actions with less movement dissimilarity across participants than complex actions and imitation actions. The error bar indicates the standard deviation.

Impact of intrinsic traits on self-recognition performance

The second analysis step focused on the impact of intrinsic variability in individual difference measures on self-action recognition. Two participants did not complete the VMIQ-2, and one participant did not complete the AQ questionnaire. Hence, listwise deletion resulted in 98 participants included in the individual differences analysis. Descriptive statistics for the scores of autistic traits (AQ score), schizotypal traits (SPQ score) and motor imagery ability (VMIQ-2 score), and subscale scores are reported in Appendix Table 1.

We first conducted regression analyses to examine relations between composite scores of individual difference measures and self-recognition performance. For each action type (simple, complex, and imitation), self-recognition performance was set as the dependent variable and three composite scores (AQ for autistic traits, SPQ scores for schizotypal traits, and VMIQ-2 scores for motor imagery ability) were predictor variables in the regression analysis. None of the three regression models revealed significant relationships. However, nonparametric correlation analyses revealed that composite motor imagery (VMIQ-2) scores negatively correlated with self-recognition performance for imitation actions (spearman $\rho = -0.221$, 95% CI [-.406, -.017], $p = .029$) (see Figure 8). Note that lower VMIQ-2 scores indicate stronger motor imagery ability. Thus, the negative correlation between motor imagery scores and self-recognition performance in imitation condition indicates that individuals with better motor imagery ability showed greater self-recognition performance for imitation actions.

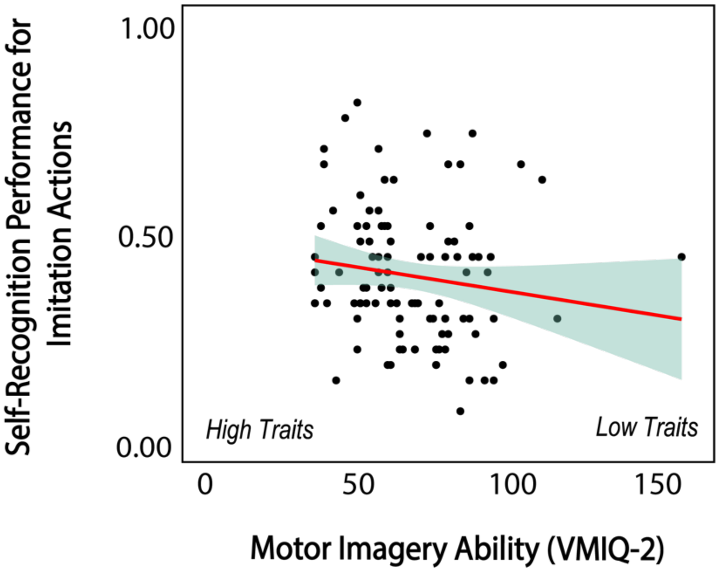


Figure 8. Relations between composite motor imagery ability and self-recognition for imitation actions and 95% confidence interval. Note that lower VMIQ-2 score indicates better motor imagery ability.

Next we examined whether subscale scores in the individual difference measures related to self-recognition performance. We focused on subscale scores for two reasons. First, both ASC and SSC are characterized by multidimensional symptom expression, which may be masked by only examining the composite measure. Second, selected subscale scores in autistic traits and schizotypal traits have been shown to relate to atypical biological motion perception (e.g., Blain et al., 2017; Puglia and Morris, 2017) as well as self-processing (e.g., Ferroni et al., 2020; Thakkar et al., 2011; Lenzenweger, 2010).

Considering the large number of possible predictors (5 subscales in AQ, 9 subscales in SPQ and 3 subscales in motor imagery VMIQ-2), we used stepwise regression on all of the possible predictors to select the important predictors in relation to self-recognition performance for the three action types. For the simple action condition, the stepwise regression selected a model with four subscale scores as strong predictor variables (AQ *social communication*, and three SPQ subscale scores including *odd behavior*, *ideas of reference,* and *no close friends*) that showed significance in predicting self-recognition performance for simple actions ($F(4,93) = 2.817$, $p = 0.030$). However, among the four selected subscale scores, only AQ *social communication* scores revealed a significant coefficient ($t(92) = -2.672$, $p = .009$). Using AQ *social communication* as the predictor, nonparametric correlation analysis revealed a negative relation between AQ *social communication* and the criterion, self-recognition for simple actions (spearman $\rho = -.229$, 95% CI [-0.414, -0.026], $p = .023$) (Figure 9). As higher AQ scores indicate more autistic traits, the negative correlation indicates that participants with more autistic traits in *social communication* were less able to recognize their actions when the actions conveyed simple types of goals.

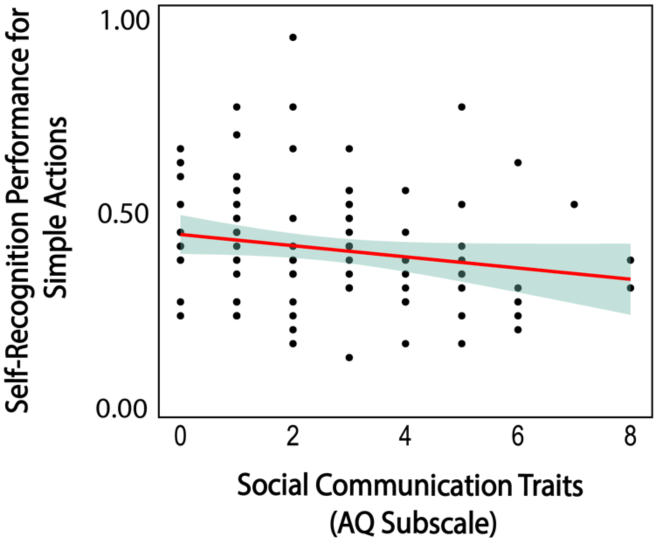


Figure 9. Relationship between AQ social communication subscale scores and self-recognition for simple actions and 95% confidence interval. Higher AQ scores indicate more autistic traits.

For complex actions with more motor planning involved in the body movements, the stepwise regression analysis selected a model with three SPQ subscale scores as predictor variables (including *unusual perceptual experiences*, *odd behavior, ideas of reference* and *no close friends)* in relation to self-recognition performance. The selected model did not reach statistical significance ($F(3,94) = 2.283$, $p = 0.084$). Upon further examination, we noticed a nonlinear quadratic relationship between SPQ *unusual perceptual experience* scores and self-recognition for complex actions. Thus, we introduced an additional predictor variable of quadratic SPQ subscale scores of *unusual perceptual experiences* to the stepwise regression analysis. The results converged to a significant model with two predictor variables (SPQ *unusual perceptual experience* subscale scores, and its quadratic term), $F(2,95) = 5.782$, $p = 0.004$. As shown in Figure 10, there was a significant relationship between the SPQ *unusual perceptual experiences* subscale and self-recognition performance for complex actions (spearman $\rho = 0.264$, 95% CI [.064, .445], $p = .008$). In addition to the linear relation, self-recognition performance for complex actions also related to a quadratic trend of SPQ *unusual perceptual* scores, revealed by significant coefficients to the quadratic term ($t(92) = -2.937$, $p = .004$). The coexistence of linear and quadratic relations indicates that participants with mid-range SPQ *unusual perceptual experiences* scores showed best self-recognition performance for complex actions.

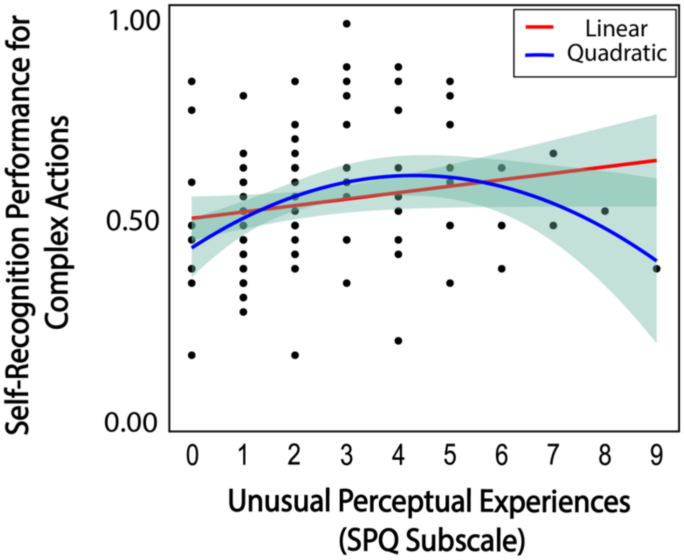


Figure 10. Relations between SPQ unusual perceptual experience subscale scores and self-recognition for complex actions and respective 95% confidence intervals. Higher SPQ scores indicate more schizotypal traits.

For imitation actions, the stepwise regression analysis selected a model with six predictor variables, including AQ *social communication* and *attention* subscale scores, three SPQ scores of *odd speech*, *suspiciousness,* and *no close friends,* and VMIQ-2 *Kinesthetic* scores, in relation to self-recognition performance ($F(6,91) = 2.895$, $p = 0.013$). However, only AQ *social communication* and *SPQ suspiciousness* scores showed significant coefficients ($t(90) = -3.286$, $p = .001$; $t(90) = - 2.336$, $p = .022$). The follow-up nonparametric correlation analysis did not reveal a significant relation between self-recognition performance for imitation actions and SPQ *suspiciousness* scores ($p = .263$), but confirmed its significant relation with the AQ *social communication* subscale scores (spearman $\rho = -0.229$, 95% CI [–0.414, -0.026], $p = .023$). This result suggests that people with higher AQ traits associated with social communication ability performed worse in the self-recognition task for imitation actions. As self-recognition performance with imitation actions also correlated with the VMIQ-2 composite scores of motor imagery ability as reported earlier, we ran a second stepwise regression analysis with four predictor variables, including AQ *social communication* scores and three VMIQ-2 subscale scores. This stepwise regression selected a model with two predictor variables, AQ *social communication* scores and VMIQ-2 *kinesthetic* motor imagery scores (*F*(2,95) = 3.652, *p* = .030). We then examined interactions between these intrinsic factors, which showed how autistic traits and motor imagery abilities jointly impact self-recognition performance for imitation actions. As shown in the right panel of Figure 11, the negative relationship between AQ *social communication* scores with self-recognition performance for imitation actions was moderated by kinesthetic motor imagery ability (*ß* = - 0.032, 95% CI [-0.063, -0.001], *p* = .043): higher *kinesthetic* motor imagery ability (- 1 *SD: VMIQ scale reversed*) was negligent in its overall effect on self-recognition (*p = .949)*, having average (*ß* = -0.034, 95% CI [-0.065, -0.027], *p* = .034) or lower *kinesthetic* motor imagery ability (+ 1 *SD*) (*ß* = -0.066, 95% CI [-0.108, -0.024], *p* = .003) significantly reduced self-recognition performance as AQ *social communication* traits increased, indicating a possible compensatory role of *kinesthetic* motor imagery ability.

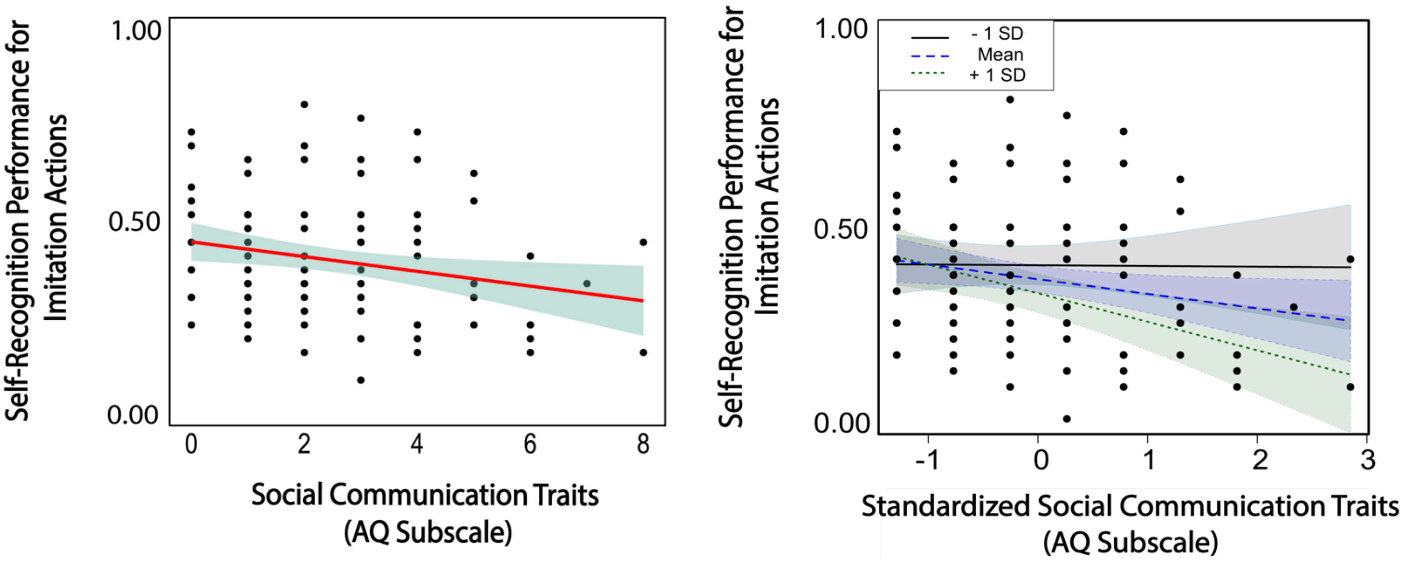


Figure 11. Relations between self-recognition for imitation actions with AQ *social communication* subscale scores (left panel) and moderation of this relationship by kinesthetic motor imagery ability (right panel). Right panel: Individual slopes were significant for those with low kinesthetic motor imagery ability (+ 1 *SD*) and average imagery ability, but not high ability (- 1 *SD*). Note that higher AQ scores indicate more autistic traits, and higher VMIQ-K scores indicate weaker imagery ability.

Discussion

On a large sample of participants, we measured the impact of motoric factors related to action planning and participant variability on self-recognition. Using quantitative measures of action similarity, we ruled out the primary contribution of movement distinctiveness to self-action recognition to show how this ability is modulated by systematic factors related to the degree of motor planning and intrinsic traits of the participant. In measuring intrinsic traits, we further parcellated the subscale measures, which revealed fine-grained variance beyond the composite measure for specific subscales. Specifically, subscales for motor imagery (kinesthetic), AQ (social communication) and SPQ (unusual perceptual experiences) interacted with joint and distinct action types to influence self-recognition performance.

Movement distinctiveness is not the primary contributor to self-action recognition in PLDs

First, we found that movement distinctiveness was not the primary contributor to variability in self-recognition performance, even after removing the visual cues to identity. Prior work has speculated on two main contributors to self-recognition of actions. On one hand, proprioceptive and kinesthetic information retained from executing familiar motor actions (i.e., motor experience) may primarily modulate self-action recognition. On the other hand, participants may recognize their movements on the basis of idiosyncratic movement cues, measured via perceptual distinctiveness (i.e., speed of movements; motion trajectories in observed actions). In our study, action dissimilarity analysis based on DTW revealed no significant relationship between perceived movement idiosyncrasies in performing actions and self-recognition accuracy. The analysis showed that even though both complex and imitation actions were significantly more idiosyncratic than simple actions and did not different from each other in terms of variability, this pattern in relation to self-recognition accuracy was distinctively different. Complex actions were self-recognized significantly better than simple and imitation actions, while simple and imitation actions showed no difference in self-recognition. Additional analyses quantifying movement speed differences across participants further found no relationship to self-recognition of the action types (see section V in Appendix), nor self-recognition differences between action types. Hence, our results suggest that neither movement variability (i.e., distinctiveness) nor low-level speed differences served as primary contributors to visual self-recognition when we accounted for other action factors.

As an additional control for visual familiarity, we added a visual recognition task for imitation actions (further detailed in Appendix). Performance on this task showed chance-level recognition of briefly observed visual actions that participants had previous visual experience with, as compared to above-chance self-recognition performance for actions that participants had previous motor experience with. Moreover, all actions, regardless of their degree of complexity, showed viewpoint invariant self-recognition (frontal and right/left half-profile) of identifying own actions, confirming a body-centered representation for self actions (Jokisch, Daüm, and Troje, 2006). Together, these findings underscore the importance of the motor system to self-recognition in that perpetual experience does not appear to suffice for self-recognition. In keeping with common coding theory, actions signaling a closer match between the motor system and perception (i.e., self-generated) enhance action recognition more so than those distally matched (i.e., other-generated) (Prinz, 1997).

Extrinsic variability modulates self-recognition performance

Our study points to alternative factors beyond visual distinctiveness that primarily underlie self-action recognition. First, we found that self-recognition performance varied according to the degree of motor planning, manipulated by our extrinsic factor of action type (simple, complex, and imitation). A few prior studies have shown the contribution of item-level delineations based on perceptual variability or idiosyncrasies to self-action recognition (e.g., Loula et al., 2005; Coste et al., 2021). Importantly, the perceptual distinctiveness of self-actions is generated from privileged access of one's motor plan during action performance (Knoblich and Prinz, 2001), and hence causally varies as a function of the degree of motor planning (e.g., Li et al., 2020). Complex actions required the greatest degree of motor planning due to the more complex variability of the goal representation space relative to simple actions (e.g., Li et al., 2020). Simple actions and imitation actions required less motor planning due to the reduced action complexity and mimicry of movement patterns, respectively (Zimmer & Engelkamp, 1996; Schwartz & Plass, 2014). The performance pattern mostly mapped onto the motor planning differences, with greatest self-recognition performance for complex actions as compared to simple or imitation actions. However, we found no performance difference between simple (simple actions with semantic label; i.e., greater motor planning) and imitation (without semantic label; i.e., reduced motor planning) actions, even though imitation actions were considerably more variable.

Intrinsic interactions with extrinsic factors influence self-recognition performance

Beyond the separate influence of extrinsic action characteristics on self-recognition, our analyses revealed distinct interactions between the extrinsic action type and intrinsic traits linked to motor imagery ability and clinical subscales (AQ social communication, and SPQ unusual perceptual experiences). For imitation actions, we found that the composite motor imagery score was significantly correlated with self-recognition performance. Motor simulation theories (Jeannerod 2001; Jeannerod and Pacherie, 2004) posit the central role of motor imagery during action perception. Our results suggest that a compensatory increase in motor imagery ability may be required to self-identify one's own imitation action, which consisted of the least amount of motor planning since no explicit action label was provided during the action execution session. No significant relationships were found between the visual imagery subtypes (internal and external) and self-recognition for any action types. This behavioral dissociation between visual and kinesthetic imagery neatly converges with neural evidence revealing partially distinct brain regions for visual imagery (parieto-occipital) and kinesthetic imagery (i.e., parietal and premotor) (Guillot et al., 2009). It is likely that during the perception of one's own action, motor expertise of own actions is a fundamental contributor, and makes use of neural resources linked to kinesthetic motor imagery (Wei and Luo, 2010; Kim et al., 2011).

We further found a negative relation between the AQ social communication subscale and self-recognition performance for imitations actions. This negative relationship was moderated by kinesthetic motor imagery traits: greater AQ social communication traits attenuated self-recognition performance for imitation actions, only when kinesthetic motor imagery ability was weakened (low or average). In other words, greater kinesthetic motor imagery ability may play a compensatory role in the Autism-Spectrum, evidenced by reduced effects of AQ social communication on self-action recognition with more kinesthetic motor imagery traits. As motor imagery is thought to share neural resources with biological motion perception (e.g., Grèzes and Decety 2001), the negative AQ social communication relation could be driven by difficulties in biological motion processing. This is consistent with neuroimaging evidence showing reduced brain activity in prefrontal and left temporal cortices in high AQ social communication individuals when viewing point-light actions (Puglia and Morris, 2017). Behaviorally, this interpretation is also in line with empirical work linking the AQ social communication subscale to atypical processing of local kinematic information (i.e., individual joint movements) in biological motion (Wang et al., 2018). For simple actions with less perceptual variability, superior local processing of joint movements may be an important factor in discriminating fine details of self from other actions, possibly compromised in those with high AQ social communication traits. This may similarly explain the nonsignificant relationship between AQ social communication and complex actions, which may be aided by the increased perceptual variability in the actions resulting from individualized body movements, providing a compensatory strategy to self-recognition, regardless of the AQ spectrum. The interpretation of weakened local processing in high AQ social communication traits is in contrast to behavioral patterns observed in biological motion research that utilize the composite AQ measure. Many of these studies show the reverse pattern—superior local processing and a generalized disturbance in global processing in high AQ individuals. However, findings from these studies remain inconsistent (largely attributed to task-specific variability) (e.g., van Boxtel et al., 2017) and a possible reconciliation could be obtained by exploring the subscales, rather than averaging them in the aggregate measure.

Scores on SPQ unusual perceptual experiences were quadratically related to complex actions, such that low and high individuals performed worse at self-recognizing complex action sequences. How might the intrinsic SPQ subscale modulate self-recognition for complex actions? Prior work suggests impaired action perception across the schizophrenia spectrum (Kim et al., 2005, Kim et al., 2011). However, we only observed a significant trend with self-recognition performance on one action type (i.e., complex actions). Thus, the present results cannot be reduced to a generalized abnormality in biological motion perception. Rather, atypical action monitoring in suppressing the sensory consequences of action may produce false positives — misattributions that the action is externally generated (Frith et al., 2000). The unusual perceptual subscale focuses on the positive (i.e., first-rank symptoms) of SSC, including incoherent perceptual experiences that reduce one’s sense of agency (Asai et al., 2008), blurring the boundaries between the self and other (e.g., Waters & Badcock, 2010; Postmes et al., 2014). In relation to atypical action monitoring, prior studies have shown that individuals high in SPQ unusual perceptual traits are more successful at self-induced tickling (i.e., Lemaitre, Luyat, & Lafargue, 2016), likened to a disrupted match between predicted efference copies and sensorimotor outcomes; see forward model accounts (Wolpert and Miall, 1996; Postmes et al., 2014). The action monitoring atypicality may be particularly exacerbated by complex actions that require a greater degree of motor planning and preparation. Weakened self-recognition performance on complex actions with those low on SPQ unusual perceptual experiences traits is less clear. It is possible that the weakened self-recognition may be due to decreased perceptual acuity in discerning subtle movement patterns, induced by the distraction of perceptual variance in complex action sequences. This could manifest in behavioral differences based on false negatives (i.e., decreased discrimination of perceptual idiosyncrasies), rather than on false positives (i.e., misattributions).

Neither composite subclinical measure (schizotypal or autistic) showed a relation to self-recognition performance. If our analysis was restricted to the composite measures, this would suggest that physical aspects of the self (i.e., bodily recognition) are intact in both conditions, consistent with many studies on visual self-recognition (e.g., Ferrari & Matthews, 1983; Uddin et al., 2008; Dunphy-Lelii & Wellman, 2012; Gillespie-Smith et al., 2014; Lee et al., 2007; cf. Chakraborty and Chakrabarti, 2015). However, the clear pattern of relations with the subscales suggests that a focus on the composite measure can be misleading, as it masks the psychopathological variance in the general population. SSC and ASC have heterogenous symptomatology, residing on a continuum that extends well into the general population, and frequently updated in light of DSM revisions. While models of AQ and SPQ symptomatology posit the large phenotypic overlap between AQ and SPQ subscales (which our results similarly maintain), the AQ social communication and SPQ unusual perceptual experiences dimensions appear to reside on a diametric axis, largely exempt from the overlap (Dinsdale et al 2013; Ford and Crewther, 2014). The present results confirm the diametric relation presented in the prior work, as we found no relationship between the subdimensions of interest (i.e., AQ social communication and SPQ unusual perceptual experiences), which shared unique variance with each respective clinical composite measure. Since the composite measures average across the spectrum, a foundational approach for future individual difference studies should also prioritize subscale parcellations, which may reveal finer-grained individual variability.

Our study presents some outstanding limitations. Though we separated actions based on motor planning, our actions included a range of social and non-social actions. In doing so, we preserve the naturalistic nature of the actions; however, additional studies can more specifically isolate the contribution of controlled action types. Another limitation pertained to the action similarity analysis in the present paper. The DTW analysis only derives similarity by comparing joint movements, not considering other factors that can influence similarity judgments such as body structure similarity across individuals, or semantic similarity across actions. In addition, DTW gives the same weight to each frame in the video. In some actions, subtle movement differences in a short period may play a more important role in determining action similarity than movements in other periods. These subtle movement segments, not capturable with dynamic time warping analysis across groups, may also be needed to assess whether motor performance differences are associated with our psychometric traits (e.g., Walker et al., 1999; Lenzenweger & Maher, 2002) or may provide idiosyncratic cues to identity in other ways. Importantly, our results do not rule out that perceptual distinctiveness is a relevant cue to self-identification, as was found in Coste et al (2021). Rather, we consider perceptual variance to be a result of the underlying motoric complexity afforded by different action goals and may be more important when other manipulations to these motoric factors are not present. As Coste et al (2021) focused on postural movements, the goal complexity remained relatively constant across action sequences, hence providing a possible explanation for why visual distinctiveness served as the primary cue.

Few species are capable of self-recognition and develop motor-based, self-other mapping mechanisms in the brain. Visual self-recognition of point-light actions provides a unique lens to gauge the core perceptual and motor mechanisms underlying self-representations in humans. These paradigms control the level of visual familiarity people have with self-related stimuli (e.g., compared to familiar faces, body images, voices) from a third-person perspective, highlighting the importance of "acting" in "seeing" the self. Our findings thus complement a diverse body of literature in self-processing, demonstrating that self-recognition is possible for stimuli even with little visual experience. Furthermore, the approach of studying individual differences provides new evidence for *how* the motor system may contribute to self-recognition— distinct aspects of motor processing tap into intrinsic and extrinsic routes, imbuing the core of self-awareness.

Acknowledgements

We thank Tabitha Safari, Nick Ichien, Aya Strauss, Lizeth Sotelo, Justin Azarian, Marian Spannowsky, Sunhee Jin, and Nazar Flome for assistance in data collection. We thank Marco Iacoboni for helpful feedback on the manuscript. We thank Joseph Burling and Steve Thurman for help with the motion capture system. This study was supported by the National Science Foundation (BCS-1655300).

References

Ahmed, A. A., & Vander Wyk, B. C. (2013). Neural processing of intentional biological motion in unaffected siblings of children with autism spectrum disorder: An fMRI study. *Brain and Cognition*. https://doi.org/10.1016/j.bandc.2013.09.007

American Psychiatric Association. (1987). Diagnostic and statistical manual of mental disorders (3rd ed., rev.).

Asai, T., Sugimori, E., & Tanno, Y. (2008). Schizotypal personality traits and prediction of one's own movements in motor control: What causes an abnormal sense of agency? *Consciousness and Cognition*, *17*(4), 1131–1142. https://doi.org/10.1016/j.concog.2008.04.004

Asperger, H. (1944). Die "autistichen Psychopathen" im Kinder- salter. Archive fur psychiatrie und Nervenkrankheiten, 117, 76- 136. (Reprinted in Autism and Asperger syndrome, by U. Frith, Ed., 1991, Cambridge, UK: Cambridge University Press).

Binet, A. (1886). La Psychologie du Raisonnement. In Alcan (Ed.). Paris

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians.[erratum appears in J Autism Dev Disord 2001 Dec;31(6):603]. *Journal of Autism & Developmental Disorders.*

Barron, D., Swami, V., Towell, T., Hutchinson, G., & Morgan, K. D. (2015). Examination of the factor structure of the schizotypal personality questionnaire among British and Trinidadian adults. *BioMed Research International*. https://doi.org/10.1155/2015/258275

Beardsworth, T., & Buckner, T. (1981). The ability to recognize oneself from a video recording of one's movements without seeing one's body. *Bulletin of the Psychonomic Society*. https://doi.org/10.3758/BF03333558

Bekkering, H., Wohlschläger, A., & Gattis, M. (2000). Imitation of gestures in children is goal-directed. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *53*(1), 153–164. https://doi.org/10.1080/713755872

Bischoff, M., Zentgraf, K., Lorey, B., Pilgramm, S., Balser, N., Baumgartner, E., Hohmann, T., Stark, R., Vaitl, D., & Munzert, J. (2012). Motor familiarity: Brain activation when watching kinematic displays of one's own movements. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2012.05.009

Blain, S. D., Peterman, J. S., & Park, S. (2017). Subtle cues missed: Impaired perception of emotion from gait in relation to schizotypy and autism spectrum traits. *Schizophrenia Research*, *183*(September 2017), 157–160. https://doi.org/10.1016/j.schres.2016.11.003

Blake, R., & Shiffrar, M. (2007). Perception of human motion. In *Annual Review of Psychology*. https://doi.org/10.1146/annurev.psych.57.102904.190152

Bläsing, B. E., & Sauzet, O. (2018). My action, my self: Recognition of self-created but visually unfamiliar dance-like actions from point-light displays. *Frontiers in Psychology*, *9*(OCT), 1–9. https://doi.org/10.3389/fpsyg.2018.01909

Bleuler, E. Dementia Praecox: Or the Group of Schizophrenias. New York: International Universities Press; 1911.

Borda, J. P., & Sass, L. A. (2015). Phenomenology and neurobiology of self disorder in schizophrenia: Primary factors. *Schizophrenia Research*, *169*(1–3), 464–473. https://doi.org/10.1016/j.schres.2015.09.024

Bralten, J., Van Hulzen, K. J., Martens, M. B., Galesloot, T. E., Arias Vasquez, A., Kiemeney, L. A., Buitelaar, J. K., Muntjewerff, J. W., Franke, B., & Poelmans, G. (2018). Autism spectrum disorders and autistic traits share genetics and biology. *Molecular Psychiatry*, *23*(5), 1205–1212. https://doi.org/10.1038/mp.2017.98

Burling, J. M., Kadambi, A., Safari, T., & Lu, H. (2019). The impact of autistic traits on self-recognition of body movements. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2018.02687

Casile, A., & Giese, M. A. (2006). Nonvisual motor training influences biological motion perception. *Current Biology*. https://doi.org/10.1016/j.cub.2005.10.071

Chakraborty, A., & Chakrabarti, B. (2015). Is it me? Self-recognition bias across sensory modalities and its relationship to autistic traits. *Molecular Autism*, *6*(1), 1–8. https://doi.org/10.1186/s13229-015-0016-1

Clark, S., Tremblay, F., & Ste-Marie, D. (2004). Differential modulation of corticospinal excitability during observation, mental imagery and imitation of hand actions. *Neuropsychologia*. https://doi.org/10.1016/S0028-3932(03)00144-1

Cohen, A. S., Matthews, R. A., Najolia, G. M., & Brown, L. A. (2010). Toward a more psychometrically sound brief measure of Schizotypal traits: Introducing the SPQ-Brief revised. *Journal of Personality Disorders*. https://doi.org/10.1521/pedi.2010.24.4.516

Conson, M., Aromino, A. R., & Trojano, L. (2010). Whose hand is this? Handedness and visual perspective modulate self/other discrimination. Experimental Brain Research, 206, 449-453.

Cook, R., Johnston, A., & Heyes, C. (2012). Self-recognition of avatar motion: How do I know it's me? *Proceedings of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rspb.2011.1264

Coste, A., Bardy, B. G., Janaqi, S., Słowiński, P., Tsaneva-Atanasova, K., Goupil, J. L., & Marin, L. (2021). Decoding identity from motion: how motor similarities colour our perception of self and others. *Psychological research*, *85*(2), 509-519.

Cutting, J. E., & Kozlowski, L. T. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society*. https://doi.org/10.3758/BF03337021

Daprati, E., & Sirigu, A. (2002). Laterality effects on motor awareness. *Neuropsychologia*. https://doi.org/10.1016/S0028-3932(01)00203-2

Daprati, E., Wriessnegger, S., & Lacquaniti, F. (2007). Knowledge of one's kinematics improves perceptual discrimination. *Consciousness and Cognition*. https://doi.org/10.1016/j.concog.2006.03.001

Dinsdale, N. L., Hurd, P. L., Wakabayashi, A., Elliot, M., & Crespi, B. J. (2013). How are autism and schizotypy related? Evidence from a non-clinical population. PloS one, 8(5), e63316.

Dunphy-Lelii, S., & Wellman, H. M. (2012). Delayed self-recognition in autism: A unique difficulty?. Research in Autism Spectrum Disorders, 6(1), 212-223.

Engelkamp, J., & Krumnacker, H. (1980). Image- and motor-processes in the retention of verbal materials. [Image- and motor-processes in the retention of verbal materials.]. *Zeitschrift Für Experimentelle Und Angewandte Psychologie*.

Ferrari, M., & Matthews, W. S. (1983). Self-recognition deficits in autism: Syndrome-specific or general developmental delay? *Journal of Autism and Developmental Disorders*, *13*(3), 317–324. https://doi.org/10.1007/BF01531569

Ferri, F., Frassinetti, F., Ardizzi, M., Costantini, M., & Gallese, V. (2012). A sensorimotor network for the bodily self. Journal of Cognitive Neuroscience, 24, 1584-1595.

Ferri, F., Frassinetti, F., Costantini, M., & Gallese, V. (2011). Motor simulation and the bodily self. PLOS ONE, 6, e17927.

Ferroni, F., Ardizzi, M., Ferri, F., Tesanovic, A., Langiulli, N., Tonna, M., Marchesi, C., & Gallese, V. (2020). Schizotypy and individual differences in peripersonal space plasticity. *Neuropsychologia*, *147*(May), 107579. https://doi.org/10.1016/j.neuropsychologia.2020.107579

Flach, R., Knoblich, G., & Prinz, W. (2004). Recognizing one's own clapping: the role of temporal cues. *Psychological Research*. https://doi.org/10.1007/s00426-003-0165-2

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. Nature Reviews Neuroscience, 10(1), 48-58.

Ford, T. C., & Crewther, D. P. (2014). Factor analysis demonstrates a common schizoidal phenotype within autistic and schizotypal tendency: Implications for neuroscientific studies. *Frontiers in Psychiatry*. https://doi.org/10.3389/fpsyt.2014.00117

Frassinetti, F., Pavani, F., Zamagni, E., Fusaroli, G., Vescovi, M., Benassi, M., ... & Farnè, A. (2009). Visual processing of moving and static self body-parts. Neuropsychologia, 47(8-9), 1988-1993.

Frassinetti, F., Ferri, F., Maini, M., Benassi, M. G., & Gallese, V. (2011). Bodily self: An implicit knowledge of what is explicitly unknown. Experimental Brain Research, 212, 153-160.

Frith, C. D., Blakemore, S. J., & Wolpert, D. M. (2000). Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. *Brain Research Reviews*. https://doi.org/10.1016/S0165-0173(99)00052-1

Gallese, V., & Ferri, F. (2014). Psychopathology of the bodily self and the brain: The case of schizophrenia. *Psychopathology*, *47*(6), 357–364. https://doi.org/10.1159/000365638

Ganesh, S., van Schie, H. T., Cross, E. S., de Lange, F. P., & Wigboldus, D. H. J. (2015). Disentangling neural processes of egocentric and allocentric mental spatial transformations using whole-body photos of self and other. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2015.05.003

Gavrila, D. M., & Davis, L. S. (1995). Towards 3-D model-based tracking and recognition of human movement: a multi-view approach. *Gesture*.

Gessaroli, E., Andreini, V., Pellegri, E., & Frassinetti, F. (2013). Self-face and self-body recognition in autism. *Research in Autism Spectrum Disorders*. https://doi.org/10.1016/j.rasd.2013.02.014

Gillespie-Smith, K., Ballantyne, C., Branigan, H. P., Turk, D. J., & Cunningham, S. J. (2018). The I in autism: Severity and social functioning in autism are related to self-processing. British Journal of Developmental Psychology, 36(1), 127–141. https://doi.org/10.1111/bjdp.12219

Grèzes, J., Fonlupt, P., Bertenthal, B., Delon-Martin, C., Segebarth, C., & Decety, J. (2001). Does perception of biological motion rely on specific brain regions? *NeuroImage*. https://doi.org/10.1006/nimg.2000.0740

Guillot, A., Collet, C., Nguyen, V. A., Malouin, F., Richards, C., & Doyon, J. (2009). Brain activity during visual versus kinesthetic imagery: An fMRI study. *Human Brain Mapping*, *30*(7), 2157–2172. https://doi.org/10.1002/hbm.20658

Hohmann, T., Troje, N. F., Olmos, A., & Munzert, J. (2011). The influence of motor expertise and motor experience on action and actor recognition. *Journal of Cognitive Psychology*. https://doi.org/10.1080/20445911.2011.525504

Hur, J. W., Blake, R., Cho, K. I. K., Kim, J., Kim, S. Y., Choi, S. H., Kang, D. H., & Kwon, J. S. (2016). Biological motion perception, brain responses, and schizotypal personality disorder. *JAMA Psychiatry*, *73*(3), 260–267. https://doi.org/10.1001/jamapsychiatry.2015.2985

Iacoboni, M., & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. In *Nature Reviews Neuroscience*. https://doi.org/10.1038/nrn2024

Jeannerod, M. (2001). Neural simulation of action: A unifying mechanism for motor cognition. *NeuroImage*. https://doi.org/10.1006/nimg.2001.0832

Jeannerod, M., & Pacherie, E. (2004). *Agency, Simulation and Self-identification*. *19*(2), 113–146.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*. https://doi.org/10.3758/BF03212378

Johns, L. C., & Van Os, J. (2001). The continuity of psychotic experiences in the general population. In *Clinical Psychology Review*. https://doi.org/10.1016/S0272-7358(01)00103-9

Jokisch, D., Daum, I., & Troje, N. F. (2006). Self recognition versus recognition of others by biological motion: Viewpoint-dependent effects. *Perception*. https://doi.org/10.1068/p5540

Kanner, L. (1943). Autistic disturbances of affective contact. Nervous child, 2(3), 217-250.

Kennel, C., Hohmann, T., & Raab, M. (2014). Action perception via auditory information: Agent identification and discrimination with complex movement sounds. *Journal of Cognitive Psychology*, *26*(2), 157–165. https://doi.org/10.1080/20445911.2013.869226

Kim, J., Doop, M. L., Blake, R., & Park, S. (2005). Impaired visual recognition of biological motion in schizophrenia. *Schizophrenia Research*. https://doi.org/10.1016/j.schres.2005.04.006

Kim, J., Park, S., & Blake, R. (2011). Perception of Biological Motion in Schizophrenia and Healthy Individuals: A Behavioral and Fmri Study. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0019971

Kraeplin, E., 1919. Dementia Praecox and Paraphrenia (R.M. Barclay, G.M. Robertson, Trans.). E and S Livingstone. Edinburgh.

Knoblich, G., & Flach, R. (2001). Predicting the effects of actions: Interactions of Perception and Action. *Psychological Science*, *12*(6), 467–472. https://doi.org/10.1111/1467-9280.00387

LaViola, J. J. (2003). Double exponential smoothing: An alternative to Kalman filter-based predictive tracking. *Proceedings of the Workshop on Virtual Environments, EGVE'03*. https://doi.org/10.1145/769953-769976

Lee, J., Kwon, J. S., Shin, Y. W., Lee, K. J., & Park, S. (2007). Visual self-recognition in patients with schizophrenia. *Schizophrenia Research*, *94*(1-3), 215-220.

Lemaitre, A. L., Luyat, M., & Lafargue, G. (2016). Individuals with pronounced schizotypal traits are particularly successful in tickling themselves. *Consciousness and Cognition*. https://doi.org/10.1016/j.concog.2016.02.005

Lenzenweger, M. F., & Maher, B. A. (2002). Psychometric schizotypy and motor performance. *Journal of Abnormal Psychology*. https://doi.org/10.1037/0021-843X.111.4.546

Li, X., Krol, M. A., Jahani, S., Boas, D. A., Tager-Flusberg, H., & Yücel, M. A. (2020). Brain correlates of motor complexity during observed and executed actions. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-67327-5

Loula, F., Prasad, S., Harber, K., & Shiffrar, M. (2005). *Recognizing People From Their Movement*. https://doi.org/10.1037/0096-1523.31.1.210

Miller, L. E., & Saygin, A. P. (2013). Individual differences in the perception of biological motion: Links to social cognition and motor imagery. *Cognition*. https://doi.org/10.1016/j.cognition.2013.03.013

Mitchell, R. W., & Curry, C. (2016). Self-Recognition and Other-Recognition in Point-Light Displays. *Open Journal of Philosophy*, *06*(01), 42–50. https://doi.org/10.4236/ojpp.2016.61005

Molnar-Szakacs, I., Kaplan, J., Greenfield, P. M., & Iacoboni, M. (2006). Observing complex action sequences: The role of the fronto-parietal mirror neuron system. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2006.07.035

Mul, C. L., Cardini, F., Stagg, S. D., Sadeghi Esfahlani, S., Kiourtsoglou, D., Cardellicchio, P., & Aspell, J. E. (2019). Altered bodily self-consciousness and peripersonal space in autism. *Autism*, *23*(8), 2055-2067.

Murgia, M., Hohmann, T., Galmonte, A., Raab, M., & Agostini, T. (2012). Recognising one's own motor actions through sound: The role of temporal factors. *Perception*, *41*(8), 976–987. https://doi.org/10.1068/p7227

Noel, J. P., Cascio, C. J., Wallace, M. T., & Park, S. (2017). The spatial self in schizophrenia and autism spectrum disorder. Schizophrenia research, 179, 8-12.

Oberman, L. M., & Ramachandran, V. S. (2007). The simulating social mind: The role of the mirror neuron system and simulation in the social and communicative deficits of autism spectrum disorders. In *Psychological Bulletin*. https://doi.org/10.1037/0033-2909.133.2.310

Parnas, J., & Henriksen, M. G. (2014). Disordered self in the schizophrenia spectrum: A clinical and research perspective. *Harvard Review of Psychiatry*, *22*(5), 251–265. https://doi.org/10.1097/HRP.0000000000000040

Perrykkad, K., & Hohwy, J. (2020). Modelling Me, Modelling You: the Autistic Self. *Review Journal of Autism and Developmental Disorders*, *7*(1). https://doi.org/10.1007/s40489-019-00173-y

Pham, C. H., Le, Q. K., & Le, T. H. (2014). Human action recognition using dynamic time warping and voting algorithm. *VNU Journal of Science: Computer Science and Communication Engineering*, *30*(3).

Postmes, L., Sno, H. N., Goedhart, S., van der Stel, J., Heering, H. D., & de Haan, L. (2014). Schizophrenia as a self-disorder due to perceptual incoherence. *Schizophrenia Research*, *152*(1), 41–50. https://doi.org/10.1016/j.schres.2013.07.027

Prasad, S., & Shiffrar, M. (2009). Viewpoint and the Recognition of People From Their Movements. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/a0012728

Prinz, W. (1997). Perception and Action Planning. *European Journal of Cognitive Psychology*. https://doi.org/10.1080/713752551

Puglia, M. H., & Morris, J. P. (2017). Neural response to biological motion in healthy adults varies as a function of autistic-like traits. *Frontiers in Neuroscience*, *11*(JUL), 1–10. https://doi.org/10.3389/fnins.2017.00404

Raine, A. (1991). The spq: A scale for the assessment of schizotypal personality based on DSM-III-r criteria. *Schizophrenia Bulletin*. https://doi.org/10.1093/schbul/17.4.555

Raine, A., Reynolds, C., Lencz, T., Scerbo, A., Triphon, N., & Kim, D. (1994). Cognitive-perceptual, interpersonal, and disorganized features of schizotypal personality. *Schizophrenia Bulletin*. https://doi.org/10.1093/schbul/20.1.191

Repp, B. H., & Knoblich, G. (2004). Perceiving action identity: How pianists recognize their own performances. *Psychological Science*. https://doi.org/10.1111/j.0956-7976.2004.00727.x

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

Roberts, R., Callow, N., Hardy, L., Markland, D., & Bringer, J. (2008). Movement imagery ability: Development and assessment of a revised version of the vividness of movement imagery questionnaire. *Journal of Sport and Exercise Psychology*. https://doi.org/10.1123/jsep.30.2.200

Ruzich, E., Allison, C., Smith, P., Watson, P., Auyeung, B., Ring, H., & Baron-Cohen, S. (2016). Subgrouping siblings of people with autism: Identifying the broader autism phenotype. *Autism Research*, *9*(6), 658–665. https://doi.org/10.1002/aur.1544

Sandsten, K. E., Nordgaard, J., Kjaer, T. W., Gallese, V., Ardizzi, M., Ferroni, F., Petersen, J., & Parnas, J. (2020). Altered self-recognition in patients with schizophrenia. *Schizophrenia Research*, *218*, 116–123. https://doi.org/10.1016/j.schres.2020.01.022

Sass, L. A., & Parnas, J. (1998). *Schizophrenia , Consciousness , and the Self.* 427–444.

Schneider, K. (1959). Clinical psychopathology. (Trans. by M. W. Hamilton) (5th ed.). Grune & Stratton.

Schwartz, R. N., & Plass, J. L. (2014). Click versus drag: User-performed tasks and the enactment effect in an interactive multimedia environment. *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2014.01.012

Sevdalis, V., & Keller, P. E. (2009). Self-recognition in the perception of actions performed in synchrony with music. *Annals of the New York Academy of Sciences*. https://doi.org/10.1111/j.1749-6632.2009.04773.x

Siever, L. J., & Davis, K. L. (2004). The Pathophysiology of Schizophrenia Disorders: Perspectives from the Spectrum. *American Journal of Psychiatry*, *161*(3), 398–413. https://doi.org/10.1176/appi.ajp.161.3.398

Thakkar, K. N., Nichols, H. S., McIntosh, L. G., & Park, S. (2011). Disturbances in body ownership in schizophrenia: Evidence from the rubber hand illusion and case study of a spontaneous out-of-body experience. *PLoS ONE*, *6*(10). https://doi.org/10.1371/journal.pone.0027089

Thurman, S. M., van Boxtel, J. J., Monti, M. M., Chiang, J. N., & Lu, H. (2016). Neural adaptation in pSTS correlates with perceptual aftereffects to biological motion and with autistic traits. Neuroimage, 136, 149-161.

Torbet, G., Schulze, D., Fiedler, A., & Reuter, B. (2015). Assessment of self-disorders in a non-clinical population: Reliability and association with schizotypy. *Psychiatry Research*. https://doi.org/10.1016/j.psychres.2015.05.011

van Boxtel, J. J. A., & Lu, H. (2013). A biological motion toolbox for reading, displaying, and manipulating motion capture data in research settings. *Journal of Vision*. https://doi.org/10.1167/13.12.7

van Boxtel, J. J. A., Peng, Y., Su, J., & Lu, H. (2017). Individual differences in high-level biological motion tasks correlate with autistic traits. *Vision Research*. https://doi.org/10.1016/j.visres.2016.11.005

Viviani, P., & Stucchi, N. (1992). Biological movements look uniform: Evidence of motor-perceptual interactions. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 603–623. https://doi.org/10.1037//0096-1523.18.3.603

Vollema, M. G., Sitskoorn, M. M., Appels, M. C. M., & Kahn, R. S. (2002). Does the Schizotypal Personality Questionnaire reflect the biological-genetic vulnerability to schizophrenia? *Schizophrenia Research*, *54*(1–2), 39–45. https://doi.org/10.1016/S0920-9964(01)00350-4

Walker, E., Lewis, N., Loewy, R., & Palyo, S. (1999). Motor dysfunction and risk for schizophrenia. *Development and Psychopathology*, 11(3), 509-523.

Wang, Y., Wang, L., Xu, Q., Liu, D., Chen, L., Troje, N. F., He, S., & Jiang, Y. (2018). Heritable aspects of biological motion perception and its covariation with autistic traits. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1714655115

Waters, F., Woodward, T., Allen, P., Aleman, A., & Sommer, I. (2012). Self-recognition deficits in schizophrenia patients with auditory hallucinations: A meta-analysis of the literature. *Schizophrenia Bulletin*, *38*(4), 741–750. https://doi.org/10.1093/schbul/sbq144

Wei, G., & Luo, J. (2010). Sport expert's motor imagery: Functional imaging of professional motor skills and simple motor skills. *Brain Research*, *1341*, 52–62. https://doi.org/10.1016/j.brainres.2009.08.014

Wohlschläger, A., Gattis, M., & Bekkering, H. (2003). Action generation and action perception in imitation: An instance of the ideomotor principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1431), 501–515. https://doi.org/10.1098/rstb.2002.1257

Wolpert, D. M., & Miall, R. C. (1996). Forward Models for Physiological Motor Control. *Neural Networks*.

Woodbury-Smith, M. R., Robinson, J., Wheelwright, S., & Baron-Cohen, S. (2005). Screening adults for Asperger Syndrome using the AQ: A preliminary study of its diagnostic validity in clinical practice. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-005-3300-7

Wuthrich, V. M., & Bates, T. C. (2006). Confirmatory factor analysis of the three-factor structure of the Schizotypal Personality Questionnaire and Chapman schizotypy scales. *Journal of Personality Assessment*, *87*(3), 292–304. https://doi.org/10.1207/s15327752jpa8703_10

Zimmer, H. D., & Engelkamp, J. (1996). Routes to Actions and their Efficacy for Remembering. *Memory*, *4*(1), 59–78. https://doi.org/10.1080/741940663

CHAPTER 3

Specialized neural mechanisms for self-recognition from whole-body movements

Abstract: Humans can identify their actions from point-light displays even when the actions are visually degraded and depicted in unfamiliar viewpoints. When presented with these visually impoverished stimuli, participants show a behavioral self-recognition advantage, indicating a preferential role of self-actions to visual awareness. To date, we lack a direct test of the neural circuitry underlying self-action recognition of the whole body. In the present study, across two sessions, we motion-captured a range of actions from 20 participants, who returned after a delay period for functional neuroimaging. Using univariate, multivariate, and connectivity analyses, we found that self-processing of own actions recruited the action observation network (AON) spanning the frontoparietal to the temporooccipital regions. The frontoparietal regions were functionally connected to the temporooccipital regions during self-processing and showed increased activity during self-recognition, while the temporooccipital regions primarily decoded domain-general identity processing. The pattern of results further indicated a degree of left-lateralization in the frontoparietal regions and largely bilateral activity in the temporooccipital regions. Together, the findings are indicative of a specialized neural circuitry for whole-body self-recognition of actions, with prioritized roles of frontoparietal regions of the AON.

Introduction

Human life is built upon our own actions. Actions we perform entail interactions with both physical and social worlds and form the building block for self-awareness. This “bodily self” is one of the most primitive aspects of the self (Jeannerod & Pacherie, 2004; Van Den Bos & Jeannerod, 2002; Blanke, 2012; Blank, 2015) and leads to advanced and reflective forms of awareness that characterize later development of a self-concept. Prominent theories connecting the self and the other often underscore the importance of the sensorimotor system in social cognition, whereby experience with our own actions is fundamental beyond the individual self, such as for grasping the meaning behind the actions of others (e.g., Gallese & Goldman, 1998; Rizzolati and Craighero, 2004; Iacoboni, 2008; Iacoboni, 2009).

Studies in visual self-recognition demonstrate a self-processing advantage during identity recognition of whole-body actions, even when the actions are visually degraded to a significant degree, as conveyed by point-light displays (PLD)s—dot motion localized to human joints that render the mechanics of the human body. These studies show compelling evidence for both an explicit self-action recognition advantage (Loula et al., 2005; Burling et al., 2019; Sevdalis & Keller, 2010, Wöllner, 2012; Beardsworth and Buckner, 1981; Knoblich & Prinz, 2001; Coste et al., 2021; Kadambi et al., under review) in comparison to recognition of visually familiar friends or strangers, as well as an implicit self-advantage across bodily stimuli (e.g., for body parts), measured by reaction time facilitation for self-actions (e.g., Frassinetti et al., 2009; Conson, Aromino & Trojano, 2010). The studies together shed light on the very intrinsic motor construction of the self that is independent of mere perceptual attributes of the stimulus. For instance, self-recognition of actions is possible even without observing visual feedback (Knoblich & Prinz, 2001), across action types varying in complexity (Burling et al., 2019; Kadambi et al., under review), viewpoints (Jokisch, Daum & Troje, 2006), speed differences (Kadambi et al., under review), as well as linked to the degree of motor expertise (e.g., in sports; Hohmann et al., 2011), and individual differences in motor imagery ability (Kadambi et al., under review). Self-recognition from PLDs also depends on the motoric goal complexity of the action, with greater self-recognition for actions that require more motor planning and complexity (Burling et al., 2019; Kadambi et al., under review; Loula et al., 2005). Together, the behavioral evidence follows that self-action recognition does not purely rely on visual signatures of the actions, but likely on motor experience generated from familiarity in performing actions that engages sensorimotor neural systems.

Is self-recognition of whole-body actions more than an awareness of visually familiar movements? While behavioral findings on self-action recognition are well-replicated, the increased recruitment of motoric systems for whole-body self-action recognition is largely speculative. Neuroimaging studies that incorporate whole-body movements often omit the self and focus on the neural processes underlying action perception performed by other individuals. These studies converge to show that visual recognition of other people engages an action-specific network attuned to the human body known as the action observation network (AON). The AON consists of both temporooccipital regions (posterior superior temporal sulcus, pSTS; extrastriate body area, EBA; fusiform gyri), regions typically studied in visual neuroscience of actions using PLDs, as well higher-level frontoparietal circuits studied in cognitive neuroscience of actions, including the inferior parietal lobe (IPL), premotor cortex, and inferior frontal cortex (IFC). In combination between these regions, the flow of information bridges action pattern recognition from visual neuroscience and higher-level action understanding in cognitive neuroscience accounts together to comprise the AON (Urgen et al., 2019; Grèzes et al., 2003).

The AON not only supports action recognition from biological motion but is also considered a critical contributor to self-recognition as it codes the fundamental basis and link between the self and the other. Single-cell neurons in the AON found first in monkeys (Di Pellegrino et al., 1992) and later humans (Mukamel et al., 2010) demonstrate visuomotor “mirroring” capabilities, in that visual observation of a similar action goal performed by another individual evokes corresponding neural activity in motor regions of oneself. On non-motoric aspects of self-recognition, including recognition of faces and voices, increased sensorimotor (mirror neuron) involvement has also been shown for the self, in comparison to familiar others or strangers (Uddin et al., 2005; Kaplan et al., 2008; Uddin et al., 2007). Moreover, the degree of neural activity is modulated by the degree of familiarity to oneself, not only within the mirror neuron network, but based on interactions with higher-level neural systems (e.g., mentalizing network) (Molnar-Szakacs & Uddin, 2013) that engage the more reflective aspects of the self.

In the literature, there are a few studies that have examined the neural mechanisms underlying self-action recognition, but these studies limit actions to simple arm movements rather than naturalistic whole-body movements, and often use tasks that do not explicitly ask participants to provide self-identity judgments. Two functional neuroimaging (fMRI) studies provide relevant evidence of increased sensorimotor involvement during self-action processing. First, Bischoff et al (2012) used implicit measures of self-recognition from PLDs constrained to four dots in the upper body. Participants in the study observed tennis swings from PLDs of themselves or others and were asked to predict the ball flight direction. They found left-lateralized frontoparietal activity eliciting stronger activation for the self than others, involving the angular gyrus in the inferior parietal lobe (IPL) and the anterior rostral medial frontal cortex. In a working memory paradigm, Wozniak et al (2021) opted to arbitrarily assign PLDs to an identity (self, friend, stranger). PLDs associated to oneself retained in memory recruited sensorimotor systems including the left middle frontal gyrus, inferior frontal gyrus (IFG), and supplementary motor area (SMA). Notably, however, because the actions were not actually performed by oneself (i.e., arbitrary association), the motor experience generated from performing the movements was not present. Other studies using videos consisted of isolated body parts of self-actions (e.g., Macuga and Frey, 2011) further confirm recruitment of frontoparietal regions during self-recognition. However, by relying on either implicit tasks in which performance can be influenced by self-judgments, tasks that lack the motor familiarity that characterize our daily motion patterns (Wozniak et al., 2021), or isolated body parts, researchers may miss an opportunity window to identify the veridical regions and networks specialized for self-identity.

In the present study, we aimed to identify the corresponding neural circuitry underlying recognition of self-actions from whole-body movements and map the contribution of the AON to self-recognition. We incorporated both featural and quantitative assessments of the action stimuli and used multivariate decoding and encoding methods to measure the full representational space of neural regions for self-recognition. Participants were recruited in a two-session experimental design. In Session 1, we motion-captured a range of actions of participants and their gender-matched close friend performed using both visual instruction (imitation) and verbal instruction (freely performed). After a delay period, participants returned in Session 2 to undergo functional neuroimaging. For each trial of scanning, participants observed motion captured PLDs of either themselves, friends, or strangers and were tasked to identify the identity of the agent.

In keeping with prior work, we hypothesized core involvement of the action observation circuitry for all identities (self, friends, or strangers). If self-actions are preferentially prioritized by sensorimotor regions, we expected increased activity in the AON for self-actions relative to non-self actions, controlling for both visual familiarity (friend) and identity (stranger). The reverse was not predicted for either of the two other identities. Thus, in order to identify a preferential circuitry for self-actions relative to others, the AON network should be involved for both self and others, generate greater activity for self than non-self, and show inherent functional connectivity.

Methods

Participants

Twenty right-handed undergraduate participants (M_{age} = 20.55, SD_{age} = 1.73, females = 12, males = 8) were recruited from the University of California, Los Angeles, and provided payment for their participation. Sample size was determined sufficiently sensitive in keeping with prior work on neuroimaging of biological motion (e.g., Saygin et al., 2004; Chang et al., 2021; Engelen et al., 2015) and self-actions in PLDs (e.g., Bischoff et al., 2012). The study was approved by the UCLA Institutional Review board. All participants were naïve to the purpose of the study. Participants had normal or corrected-to-normal vision and no physical disabilities.

Apparatus

The Microsoft Kinect V2.0 and Kinect SDK were used for motion-capture of actions. The Kinect is effective for motion-capture and well-utilized in previous studies on self-action recognition (Kadambi et al., under review; Burling et al., 2018). Customized software developed in our lab was used to enhance movement signals, and to carry out additional processing and trimming for actions presented later in the testing phase (Van Boxtel & Lu, 2013). Three-dimensional (X-Y-Z) coordinates of the key joints were extracted at a rate of approximately 33 frames per second. Each action was trimmed to the start and stop of a T-position signaled by the participant and normalized to scale for use in the experimental task. Note that while motion capture accuracy was high, the

Kinect occasionally produced noise jittering in the stimuli, where frame-to-frame joints positions occasionally showed sudden jumps in position. Hence, to account for noisy frame-to-frame jitter, we impinged a manual correction for certain frames (i.e., replacing with the closest previous frame where the jitter was not present).

Stimuli

Twelve actions were selected from our previous work on self-action recognition (Burling, Kadambi, Safari, Lu, 2018; Kadambi and Lu, 2019; Kadambi, Xie and Lu, under review). These actions conveyed a range of variability in terms of action planning. Six of the actions (i.e. *argue, wash windows, get attention, hurry up, stretch, and play guitar)* were categorized as "verbally instructed actions", delineated by a high degree of motoric goal complexity as defined in our previous work (Burling et al., 2019; Kadambi et al., under review). These actions were verbally instructed to the participant (e.g., please perform the action: "to argue"). The remaining six actions were characterized as visually instructed (imitation) actions, depicting a range of simple and complex goals (i.e., *jumping jacks, basketball, digging, chopping, laughing, directing traffic*). For these actions, participants observed a stick figure performing an action without any verbal label provided and were then visually instructed to 'imitate the movements of the action.' These stick figure actions were selected from the Carnegie Mellon Graphics (CMU) Lab Motion Capture Database available online (http://mocap.cs.cmu.edu), generated from pre-defined actors. PLDs were thus created using the above method for each participant, a gender-matched friend, and a gender-matched stranger. The stranger action was randomly selected from one of three possible distractors for each gender (six total), pre-captured from the experimenters' and research assistants' actions. The categorization of the action types, in addition to providing variability of the action goal, further allowed us to explore secondary analyses contrasting actions involving less motor planning (visual instruction) versus actions that required a greater degree of motor planning (verbal instruction) (detailed in Appendix).

Procedure

Behavioral Session

In the first session (behavioral), participants' body movements were recorded using the Microsoft Kinect V2.0 and Kinect SDK in a quiet testing room. Here, participants were instructed to perform the actions in a rectangular space, in order to provide flexibility to perform the action, while remaining within recording distance. The Kinect was placed 1.5 m above the floor and 2.59 m away from the participant. Participants were instructed to freely perform 12 actions as described above and recorded by our motion capture system. Participants signaled the start and stop of action performance with a T-Pose. Participant actions were then recorded and converted to point-light stimuli (*Figure 12*) for use in the fMRI session. Each of the 20 participants also brought a gender-matched close friend, who was also separately recorded with the same paradigm. None of the participants were informed about the study's purpose on self-recognition, but were informed that this study was about general visual action processing. We utilized the recordings of the close friend in the fMRI session to assess the impact of visual familiarity. After the recording session, participants completed a few attitudinal

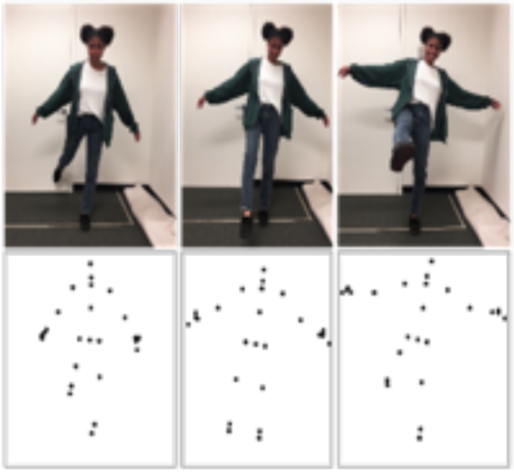


Figure 12. Conversion of participant's raw actions (kicking) into point-light displays

questionnaires including the Autism-Spectrum Quotient (Baron-Cohen et al., 2001), Schizotypal Personality Questionnaire (Raine, 1991), and Vividness of Motor Imagery-2 (Roberts, 2008).

fMRI session

Returning after a delay period of around two weeks (mean delay days = 18.55, *SD* = 2.87), participants underwent fMRI scanning in Session 2. During each trial of scanning, participants passively observed a point-light display (consisting of 25 joints) of either their own action, gender-matched familiar friend, or gender-matched stranger for a 5 second duration. Following passive observation, participants were prompted to identify with a finger-press response on button box with the right hand, whether the action video shown was their own, friend, or stranger within a 2 second maximum response period. Their response was followed by a jittered intertrial intervals (ITI) mean-centered at 5 seconds. Response order mapping of self/friend/stranger was counterbalanced across participants to reduce any effects of motor preparation or planning demands. A sample trial structure is represented in Figure 13. There were four runs per participant, each consisting of 36 trials (12 trials per identity condition) in an event-related design. For each run, experimental conditions were pseudorandomized to reduce stimulus autocorrelation related to order and sequence effects as well as correlated noise, such as scanner drift. Duration of the experimental task during scanning lasted around 24 minutes. Total scanning duration lasted around 45 minutes.

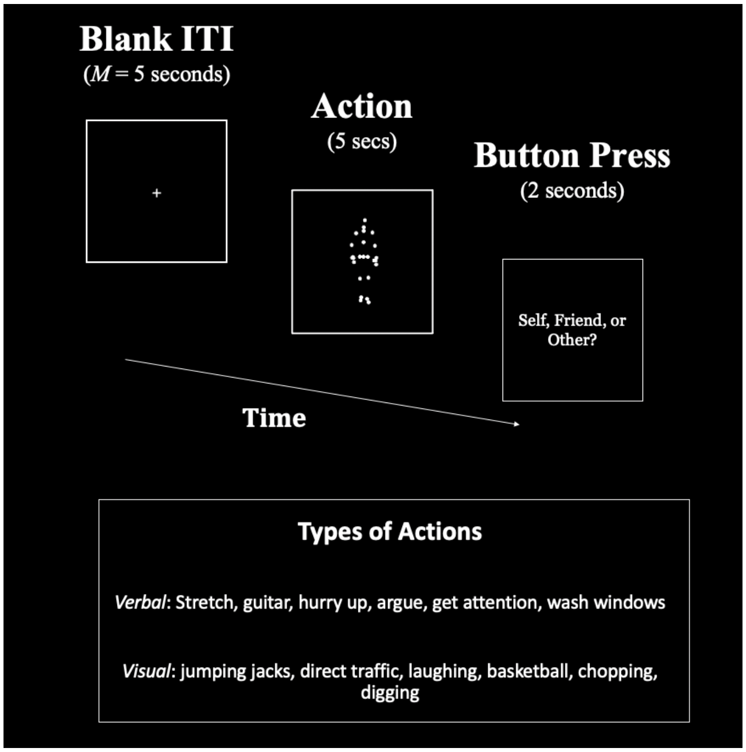


Fig 13. Trial structure including timing. Participants centrally attended to a white fixation cross until the action (self/friend/other) appeared for 5 seconds. Participants were provided 2 seconds to make their identity judgment, followed by the variable ITI. The response order of self, friend, other varied per participant in order to reduce any impact of motor order.

Localizer Sessions

Two functional localizer scans were collected for posterior superior temporal sulcus (pSTS); Grossman et al., 2000; Grossman et al., 2010) and extrastriate body area (EBA; Peelen & Downing, 2005; Downing et al., 2006; Downing et al., 2001). For the pSTS localizer, participants observed PLDs comprised of 12 dots, globally constructed to perform 12 different everyday actions (e.g., jumping, kicking, running, throwing). They also observed spatially scrambled PLDs in which the individual local dot trajectories of the point-light display remained intact, but the global configuration of the point lights was randomly displaced (i.e., starting positions of the dots were randomized). Sensitivity to intact biological motion was contrasted with spatially scrambled biological motion for the task contrast of interest. For the EBA localizer, 20 images of gray-scale bodies or chairs were presented sequentially for 300 ms. All images were identical in size (400 x

400). To ensure that participants were attending to both localizer tasks, participants were asked to focus on a central fixation cross and press a button when repetition of any stimulus occurred (i.e., one-back).

MRI Acquisition

The Siemens 3-Tesla Prisma Fit scanner at the Staglin IMHRO Center for Cognitive Neuroscience was used for Magnetic resonance imaging, equipped with a 32-channel head coil. Structural data was acquired using a T1-weighted MPRAGE protocol (2.5 mm^3 resolution; repetition time = 2.0 ms). Functional data was acquired utilizing T2*-weighted Gradient Recall Echo sequence. Scanning parameters for the main task included: repetition time = 700 ms, echo time = 33 ms, voxel size = 2.5 mm^3 isotropic voxels, field of view = 192 mm, flip angle = 70°. Four dummy scans were acquired and discarded before each scan to account for scanner stabilization. Scanning parameters for the localizer task included: repetition time = 2000 ms, echo time = 2.52 ms, voxel size = 1.0 mm3 isotropic voxels. Five dummy scans were acquired and discarded for the localizer tasks. Participants viewed the stimuli presented on a projector through a mirror mounted on the head cover in the scanner. Participants underwent four runs of 36 trials each. Each run lasted approximately 360 seconds.

Imaging Analyses

Univariate Analysis

Statistical analyses were conducted using FEAT (FMRI Expert Analysis Tool) Version 6.00, part of FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl) using the GLM approach. Individual functional scans were coregistered to the high resolution structural image using boundary-based registration (Greve and Fischl, 2009). Registration of the high resolution structural to the Montreal Neurological Institute (MNI) template was implemented using FSL's FLIRT (Jenkinson 2001, 2002) with 12 parameter DOF affine transformation. The following pre-statistics were applied; motion correction using MCFLIRT (Jenkinson 2002); slice-timing correction using Fourier-space time-series phase-shifting; non-brain removal using BET (Smith 2002); spatial smoothing using a Gaussian kernel of FWHM 5mm; grand-mean intensity normalisation of the entire 4D dataset by a single multiplicative factor; highpass temporal filtering (Gaussian-weighted least-squares straight line fitting, with sigma=50.0s). Regressors were defined based on the onsets and durations of the three identities (self, friend, stranger) across all actions. Individual runs were aggregated into a mixed effects higher-level model using FLAME (FMRIB's Local Analysis of Mixed Effects) stage 1 and stage 2 (Beckmann et al., 2003; Woolrich, 2004; Woolrich 2008) for both within-session single subject variance and between-session group level variance. Significance testing on the statistical parametric maps was then assessed at the group level using two approaches in FSL (1) *randomise* with threshold-free cluster enhancement (TFCE) cluster correction (Winkler et al., 2014; Smith & Nichols, 2009), TFCE-p threshold = .05 (2) random-field (RFT) based thresholding at $Z > 3.1$, cluster corrected to a significance level of $p < .05$ (Worsley 2001). *Randomise* served as our main approach to significance testing given its more conservative, specific, and sensitive significance criteria (Smith & Nichols, 2009). For simplicity, all figures and tables generated from the parametric RFT analysis are reported in Appendix. Conjunction analysis to localize self-specific activity was also implemented in FSL using the easythresh_conj script (easythresh_conj) on univariate activation maps for both "self vs stranger" and "self vs friend" contrasts (Nichols et al., 2005; Price & Friston, 1997). The conjunction specifically tests the "conjunction null hypothesis" as to whether both conditions showed significant functional activation ($Z > 3.1$, $p < .05$), which were later used as seed regions in the connectivity analyses.

Functional localizer

For the extrastriate body area (EBA), we measured the functional activation from the task contrast *bodies* > *chairs* elicited by the functional localizer, uncorrected $p < .05$. Given the widespread activity evoked by the contrast, we constrained activity to the anatomical parcellation of the inferior lateral occipital cortex (LOC) from the Harvard-Cortical atlas generated by FSLEYES (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases) and thresholded the mask to include the top 80% of voxels. The identical procedure was used for the posterior superior temporal sulcus (pSTS), with functional activation of interest measured by the contrast of *intact* > *scrambled* PLD and then constrained to the anatomical parcellation of the posterior superior temporal gyrus from the Harvard-Cortical Atlas. We could not localize the EBA for one participant, and the pSTS for 4 participants using this method, so for those participants, their functional ROI was replaced using the peak group level activity for each contrast with the identical method.

Functional connectivity: Psychophysiological Interaction (PPI; Friston et al., 1997)

To identify a self-action circuitry based on task-specific changes in functional connectivity, we implemented generalized PPI (gPPI) connectivity analyses (McLaren et al., 2012). PPI relies on a moderation-based GLM framework, which examines how the relationships between a seed (physiological) region and voxels in other brain regions are modulated by the psychological state of the participant (task-dependent). The degree to which the seed regions and sink (other brain regions) vary as a function of the task, is measured by testing the significance of the β coefficient of the interaction computed between the experimental contrast vector and the sink region. gPPI similarly follows this logic, but is considered more sensitive and specific as a method as it accounts for residual variance related to all experimental conditions (Cisler, Bush, & Steele, 2014). Thus, in addition to the physiological seed, the GLM for gPPI further included all task-conditions and their PPI interaction regressors. As our analyses focused on a self-action circuitry, we constrained our physiological seeds to those determined by group-level functional activations in separate GLMs for either *self > stranger* or *self > friend* contrasts. The seed region in the Left IPL was generated from creating a small sphere (2mm radius) around the peak functional activation for the conjunction of the prior contrasts. We used a conjunction analysis implemented in FSL using the easythresh_conj script (easythresh_conj) on univariate activation maps for *both* "self vs stranger" and "self vs friend" contrasts. The seed region in the Left Insula/IFG was similarly generated from univariate activation, but only from the *self > stranger* since no activation was present at the group-level for *self > friend* using the *randomise* approach. We extracted the time courses in these seed regions (Left IPL or Left Insula/IFG) using fslmeants (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Fslutils), which generated a vector of mean activity in the mask for each volume. This time course was then entered as the ROI time series regressor into the gPPI GLM. Thus, the full GLM consisted of the interaction vectors (gPPI regressors), the main effects of the contrasts of interest (the psychological variables), and a vector representing the seed region time course (the physiological variable, Y regressor). At the group level, statistical parametric maps for the interaction term were thresholded ($p < .001$, uncorrected) to compute significance of the interaction term (i.e., task modulation of interest).

Multivariate Analysis (MVPA)

Both region of interest (ROI) and whole-brain searchlight representational dissimilarity analyses (RDA) were implemented using the CoSMoMVPA toolbox (http://www.cosmomvpa.org/; Oosterhof et al., 2016) in MATLAB R2020a. Regressors were defined based on the onsets and durations of the three experimental conditions (self-actions, friend-actions, or stranger-actions). Using the Least-Squares Separate approach, beta-series parameter estimates (Rissman, Gazzaley, & D'Esposito, 2004; Mumford et al., 2012) were iteratively estimated per trial by modeling a regressor for the event of interest in the trial and a regressor for all other events within the run. Standard motion parameters were also included as regressors in each GLM. Preprocessing was identical to the univariate analysis, but no smoothing was applied. For each run, we extracted the 36 beta weights from each participant, normalized each beta weight within run, computed the average for each of the 36 action targets across all runs, and then demeaned the data (i.e., subtraction of the grand mean of all averaged targets from each averaged target). For ROI-MVPA, a linear support vector machine (SVM) was trained on neural activity patterns from three runs and tested on the remaining one run using the leave-one-run-out cross-validation measure for each participant. The same beta weights extracted for each participant were used for both ROI-MVPA and RDA.

ROI -MVPA

Feature selection

ROIs for MVPA were constrained to spherical regions implicated in bodily self-processing and action processing based on peak coordinates extracted from meta-analytic activations in the frontoparietal regions (van Veluw & Chance, 2013; Qin et al., 2020) or the temporooccipital functional localizers (EBA and pSTS). The meta-analytic regions included the bilateral inferior parietal lobule (*x,y,z* = +/- 50, -40, 34), bilateral insula (*x,y,z* = +/- 44, -2, 0), bilateral temporoparietal junction (TPJ) (Left *x,y,z*: -48, -38, 36; Right *x,y,z*: 52, -26, 38). We accounted for individual variability in the peak ROI location by defining a sphere of 3-mm radius centered around the meta-analytic coordinate and moved the sphere within 4 voxels in each *x,y,z* direction for each participant as a ROI selection step. Within this search space, we used the peak ROI at the group-level participants that distinguished self from others (*self* > *friend* and *self* > *stranger*). As the meta-analytic coordinates did not include the IFG, we also included bilateral ROIs for the probabilistic map for the IFG *pars opercularis* from the Harvard Cortical Atlas.

Table 1: Frontoparietal Locations for ROI-MVPA

Location	Hemisphere	Area	MNI(x,y,z)	Original Coordinates
Meta-Analysis	Left	IPL	(-50, -40, 26)	(-50, -40, 34)
Meta-Analysis	Right	IPL	(52, -44, 32)	(50, -40, 34)
Meta-Analysis	Left	Insula	(-46, 0, 2)	(-44, -2, 0)
Meta-Analysis	Right	Insula	(42, 6, 2)	(44, -2, 0)
Meta-Analysis	Left	TPJ	(-54, -32, 42)	(-48, -38, 36)
Meta-Analysis	Right	TPJ	(50, -24, 42)	(52, -25, 38)
Control	Left	Cerebellum	(-18, -70, -54)	(-48, -38, 36)

Abbreviations: Inferior Parietal Lobe (IPL); Temporo-parietal Junction (TPJ)

For the ROIs generated from the functional localizer scans, we extracted activity in the bilateral posterior superior temporal sulcus (pSTS) and bilateral extrastriate body area (EBA) using the method detailed in the functional localizer section. Since we encountered difficulty for some participants (*N* = 4) in extracting the pSTS localizer, we also included the full probabilistic map of the bilateral posterior superior temporal gyrus to test as ROIs.

Table 2: Temporooccipital Locations for ROI-MVPA

Location	Hemisphere	Area
Atlas	Left	IFG (pars opercularis)
Atlas	Right	IFG (pars opercularis)
Atlas	Left	pSTG
Atlas	Right	pSTG
Localizer	Left	pSTS
Localizer	Right	pSTS
Localizer	Left	EBA
Localizer	Right	EBA

Abbreviations: Inferior Frontal Gyrus (IFG), Posterior Superior Temporal Gyrus (pSTG); Posterior Superior Temporal Sulcus (pSTS); Extrastriate Body Area (EBA)

MVPA Classification

For ROI-MVPA, we conducted two-class classification to discriminate between each identity: self vs stranger; self vs friend; friend vs stranger. Classification accuracies for each ROI were compared to chance decoding (50%) using a two-tailed one-sample t-test and corrected for multiple comparisons using the false discovery rate ($q < .05$) across all participants (Benjamini and Yekutieli, 2001; Genovese, Lazar, Nichols, 2002; Pereira & Botvinick, 2011).

Representational Dissimilarity Analyses (RDA)

RDA (Haxby et al., 2014; Krieskegorte et al., 2008) was implemented using the CoSMoMVPA toolbox. We generated multiple target representational dissimilarity matrices (RDM)s based on spatiotemporal movement distinctiveness (dynamic time warping), speed, acceleration, body structure consisting of limb segment length, and a theoretical RDM based on visuomotor identity. The neural RDM was then constructed through extracting the 36 beta weights from each participant, normalizing each beta weight within run, computing the average for each of the 36 action targets across all runs, and then demeaning the data (i.e., subtraction of the grand mean of all averaged targets from each averaged target). We then computed their pairwise dissimilarity (1 – Pearson correlation) across all runs. The construction of all RDMs (behavioral, theoretical, and neural) were thus square, symmetric, and reflected the pairwise dissimilarity between each element in the matrix. For whole-brain searchlight RDA, we defined each searchlight window by a Gaussian sphere of 2mm radius. Each spherical searchlight included every voxel in the brain, along with neighboring voxels within the window. RDA was implemented through correlating the target RDM with neural RDM in each searchlight across the whole-brain. The correlations were then Fisher-z transformed and mapped to the center of each searchlight to create individual similarity maps in native space as inputs to the higher-level non-parametric analyses. Individual maps were normalized to the MNI-152 template using FSL's FLIRT functionality (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FLIRT) using trilinear interpolation for group-analysis. One-sample t-tests were computed at the group level, correcting for multiple comparisons using permutation-based threshold-free cluster enhancement with a corrected threshold of $p < .05$ (Smith and Nichols, 2009) with 10000 Monte Carlo Simulations.

Target RDMs:

Below we detail three target RDMs used for RDA. Detailed computations related to other target RDMs measured (speed and acceleration) are in the Appendix.

Movement Distinctiveness:

We generated the behavioral RDM for movement distinctiveness through implementing the dynamic time warping (DTW) algorithm to compare trajectory differences. DTW measures the pairwise movement dissimilarity between action time series via an alignment procedure that accounts for variability in time series length or duration. DTW aims to find the lowest cost function (warping path) between pairwise action time series that stretches or shrinks the time series to reflect warped distances. Greater DTW values indicate greater movement dissimilarity between time series. A 36 x 36 RDM was thus created for each participant that computed the pairwise DTW dissimilarity between each of the 12 actions across each identity (self, friend, stranger). For a full description of the DTW procedure, see Appendix materials.

Postural Limb Length:

Limb length was calculated by computing the 3D Euclidean distance between pairs of joints that made up each of the 24 limbs of the PLD. Pairwise absolute value dissimilarities were then calculated across participants for each limb and averaged together across all limbs to comprise the target RDM.

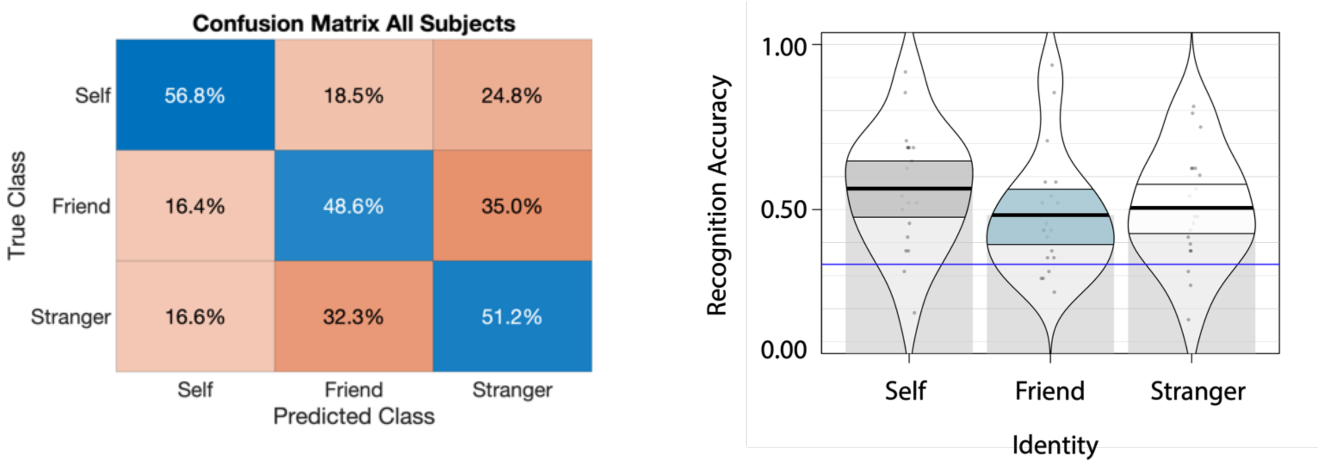
Visuomotor Identity:

We computed a theoretical RDM based on the idealized visuomotor familiarity between each of the identities. Identity for the self was coded as 0 (most familiar due to motor experience), friend was coded as 0.5 (based on the theorized overlap of common visuomotor codes), stranger was coded as 1 for all actions. Dissimilarity was computed between identities to comprise the theoretical RDM.

Results

Behavioral Analysis

Across all actions, participants could discriminate all PLD identities (self, friend, stranger) significantly above chance (.33) performance, self: *M* = .563, *SD* = .180, *t*(19) = 5.789, *p* < .001, cohen's *d* = 1.29; friend: *M* = .483, *SD* = .182, *t*(19) = 3.754, *p* = .001, *d* = .839; stranger: *M* = .5052, *SD* = .172, *t*(19) = 4.554, *p* < .001, *d* = 1.01.


Confusion Matrix All Subjects

True Class \ Predicted Class	Self	Friend	Stranger
Self	56.8%	18.5%	24.8%
Friend	16.4%	48.6%	35.0%
Stranger	16.6%	32.3%	51.2%

Fig 14. *Left panel*: depicts confusion matrix for each identity. No significant misattributions for found for the self across other identities, though friend and stranger were more confused relative to the self (~55% increase in misattributions for friend and strangers). *Right panel*: average recognition accuracy for each identity. Light gray fill indicates bar plots. Inference bands denote 95% Bayesian highest density interval with 1000 iterations. Horizontal blue line indicates chance-level decoding accuracy (.33); * *p* < .05, ** *p* ≤ .01, *** *p* ≤ .001

Recognition accuracy for self-generated actions was significantly greater than friends' actions, *t*(19) = 2.67, *p* = .015, *d* = .598, but not for strangers *t*(19) = 1.353, *p* = .192. No difference in accuracy was observed between recognition of friends versus strangers, *t*(19) = -.454, *p* = .655. No significant difference was observed between recognition for self vs friends' actions (*t*(19) = .367, *p* = .718). Greatest self-recognition accuracy was found for the stretch action (*M* = .788, *SD* =

.412) and lowest for digging *(M* = .375, *SD* = .487) (Fig 15). Across all actions, no relationships were found between self-recognition accuracy and distinctiveness related to speed (*p* = .747), acceleration (*p* = .380), postural length (*p* =.410), or spatiotemporal dissimilarity (*p* = .174)

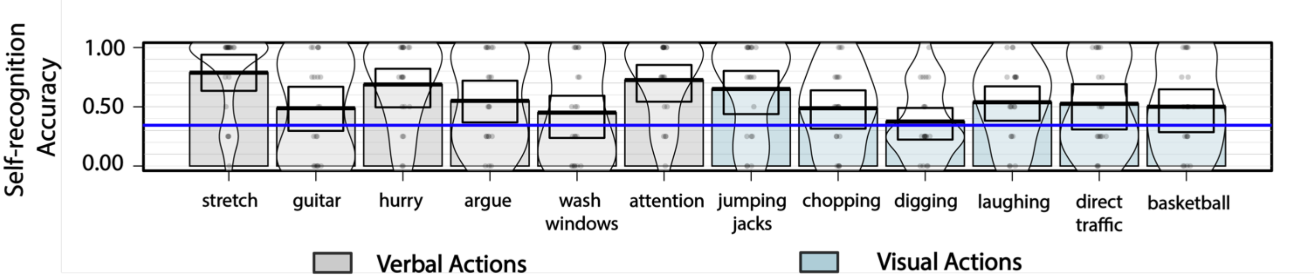


Fig 15. Self-recognition performance for different actions color coded by action type (verbal instruction: gray; visual instruction: blue). Light gray fill indicates bar plots for verbal instruction. Light blue fill indicates bar plot for visual instruction. Inference bands denote 95% Bayesian highest density interval with 1000 iterations. Horizontal blue line indicates chance-level decoding accuracy (.33)

Univariate Results

Across all identities (s*elf, friend, stranger*) vs baseline, univariate analysis revealed recruitment of the action observation network (overlayed in MNI space *Fig 16*), including the bilateral SMA (Right *x,y,z* = 12, 6, 56; Left *x,y,z* = -4, -8, 52), premotor cortex (Right *x,y,z* = 39,1,53*;* Left *x, y, z* = -45, 2, 50), inferior frontal gyri (IFG) (Right *x,y,z* = 50, 15, 10; Left *x,y,z* = -55, 16, 10), inferior parietal lobe (IPL) (Right *x,y,z* = 50, -40, 14; Left *x,y,z*: -56, -44, 11), posterior superior temporal sulcus (pSTS) (Right: *x,y,z* = 56, -42, 10, Left: -52, -50, 10) and lateral occipital cortices, spanning the extrastriate body area (EBA) (Right *x,y,z* = 44, -60, 10*; Left x,y,z,* =-51, -69, 10*).*$^{+}$

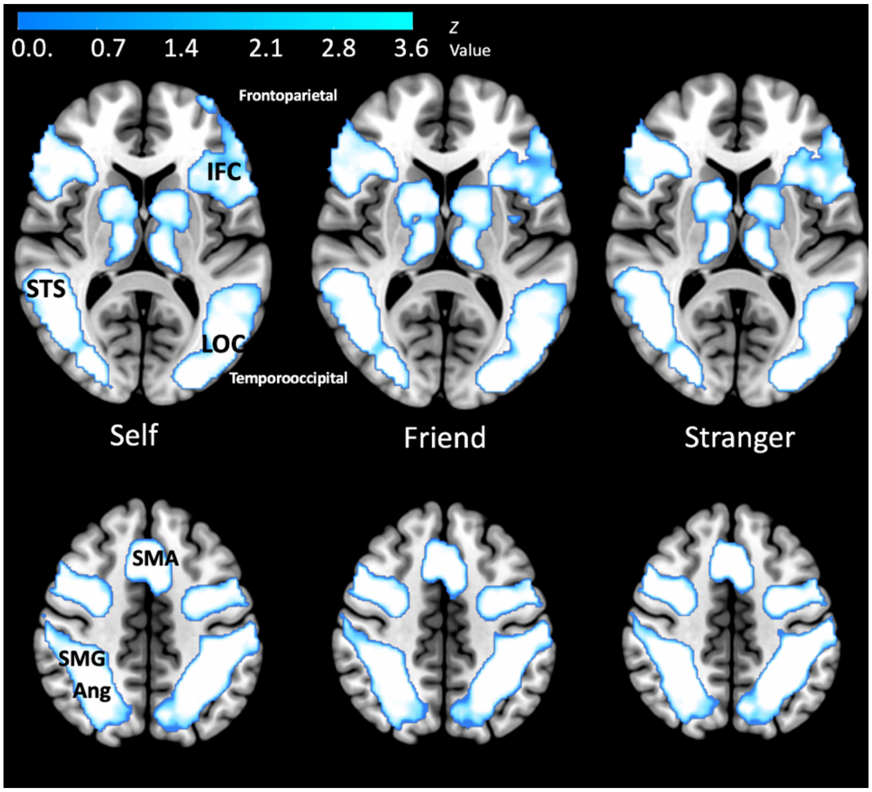


Fig 16. Group-level activity obtained using FSL's non-parametric permutation approach (*randomise*) with TFCE, $p < .05$. *From Left to Right:* Self v baseline; friend v baseline; and stranger v baseline. $^+$Large cluster sizes were obtained with TFCE due to the optimal cluster-defining threshold; hence cluster peaks are reported with visual interpolation using manual thresholding. Abbreviations: Inferior Frontal Cortex (IFC); Superior Temporal Sulcus (STS); Lateral Occipital Cortex (LOC); Supplementary Motor Area (SMA); Supramarginal Gyrus (SMG); Angular Gyrus (Ang)

The main contrasts of interest for the self (*self > stranger* and *self > friend*) similarly recruited frontoparietal regions, lateralized to the left. For *self > stranger*, activity was localized to the left posterior supramarginal gyrus (peak *x,y,z* = -62, -48, 28) into the angular gyrus, as well as spanning the left insular cortex to the inferior frontal cortex (*x,y,z* = -42, 10, -8) *(Figure 17*, left panel). Additionally, a few small clusters in the anterior cingulate cortex (ACC) (*x,y,z* = -2, 20, 18; *x,y,z* = 4, 14, 28) and one small cluster in the right insular cortex (*x,y,z* = 40, 10 ,-2) were also observed. *Self > friend* similarly recruited the left posterior SMG of the IPL (*x,y,z* = -54, -50, 30), and into the angular gyrus (*Figure 17*, right panel). Parametric RFT cluster-corrected activity further confirmed the presence of the left IFG (*x,y,z* = -47, 11, 4) for *self* > *friend* (detailed in Appendix). For *friend* > *stranger,* FSL's *randomise* approach did not yield significant activity, though parametric cluster correction ($Z > 2.3$, $p < .01$) revealed activity in the right pSTS and the dorsomedial prefrontal cortex. No significant clusters were found for either method for the contrast of *stranger* > *friend* condition. All peak clusters are reported in Appendix.

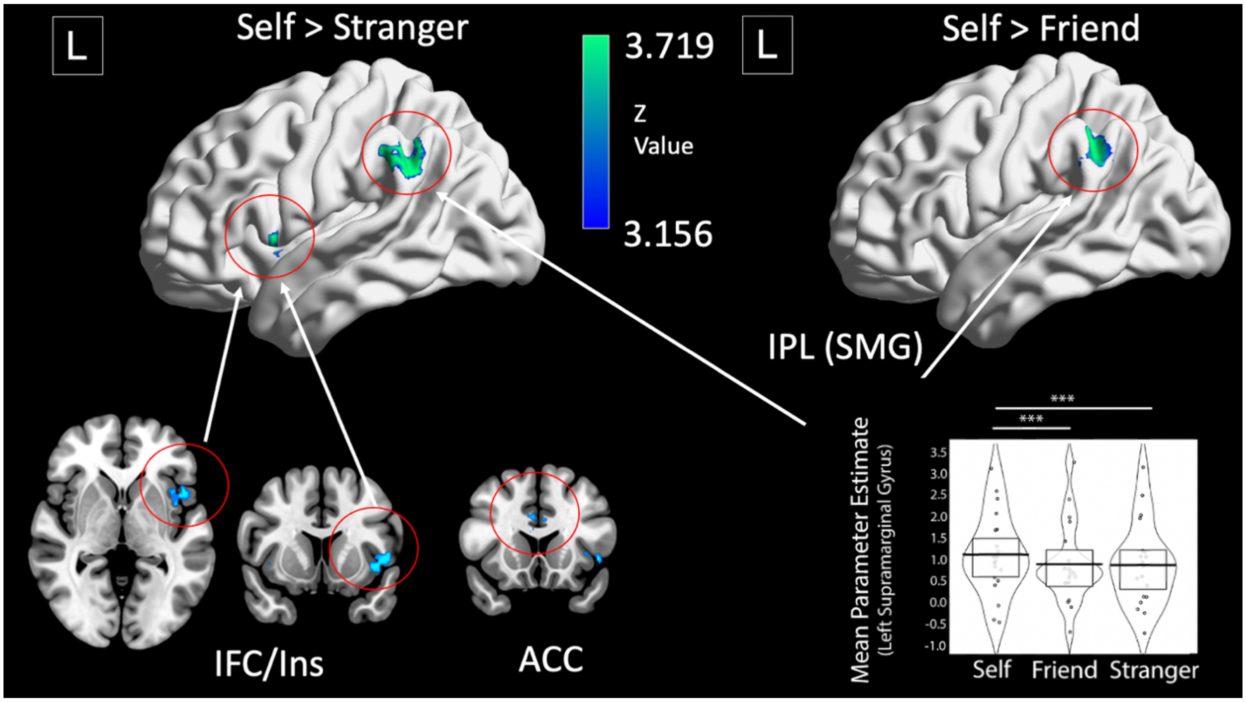


Fig 17. Univariate group-level activity ($N = 20$) for self > stranger (left) and self > friend (right) using the FSL *randomise* permutation approach, cluster corrected with TFCE ($p < .05$). Violin plot shows mean parameter estimates (PE) for the left posterior supramarginal gyrus (SMG) for all identities. The left SMG significantly discriminated contrasts of PE for both self vs stranger ($p = .001$) and self vs friend ($p = .005$), but not friend vs stranger ($p = .821$)

Results for gPPI:

Task-specific modulation during *self* > *stranger* revealed greater connectivity strength between the left IPL seed taken from the functional activation of conjunction between self > friend and self

>stranger (2*mm* sphere; centered at peak center-of-gravity *x, y, z* = -56, -44, 42) and left temporooccipital regions, extending from the middle and inferior temporo-occipital gyrus(*x, y, z* = -54, -54, -10) to the inferior lateral-occipital cortex (*x ,y, z* = -52,-70,-14), cluster corrected ($Z >$ 2.3, $p < .05$). At the more liberal threshold ($p < .001$ uncorrected), the left IPL showed a whole-brain increase in connectivity with the bilateral inferior frontal cortices (*x, y, z* = -54, 16, 30; *x, y, z* = 48, 14, 20). Similarly, task-modulation during *self* > *friend* ($p < .001$ uncorrected) of the seed left IPL also showed greater connectivity with the left inferior frontal cortex (IFC), IFG pars opercularis (*x,y,z* = -54, 20, 10) and near the caudate (*x,y,z* = 20, -20, -26). Clusters with 10 or less voxels were not reported. The IPL seed showed no seed-to-whole-brain increase in connectivity during *friend* > *stranger* contrasts.

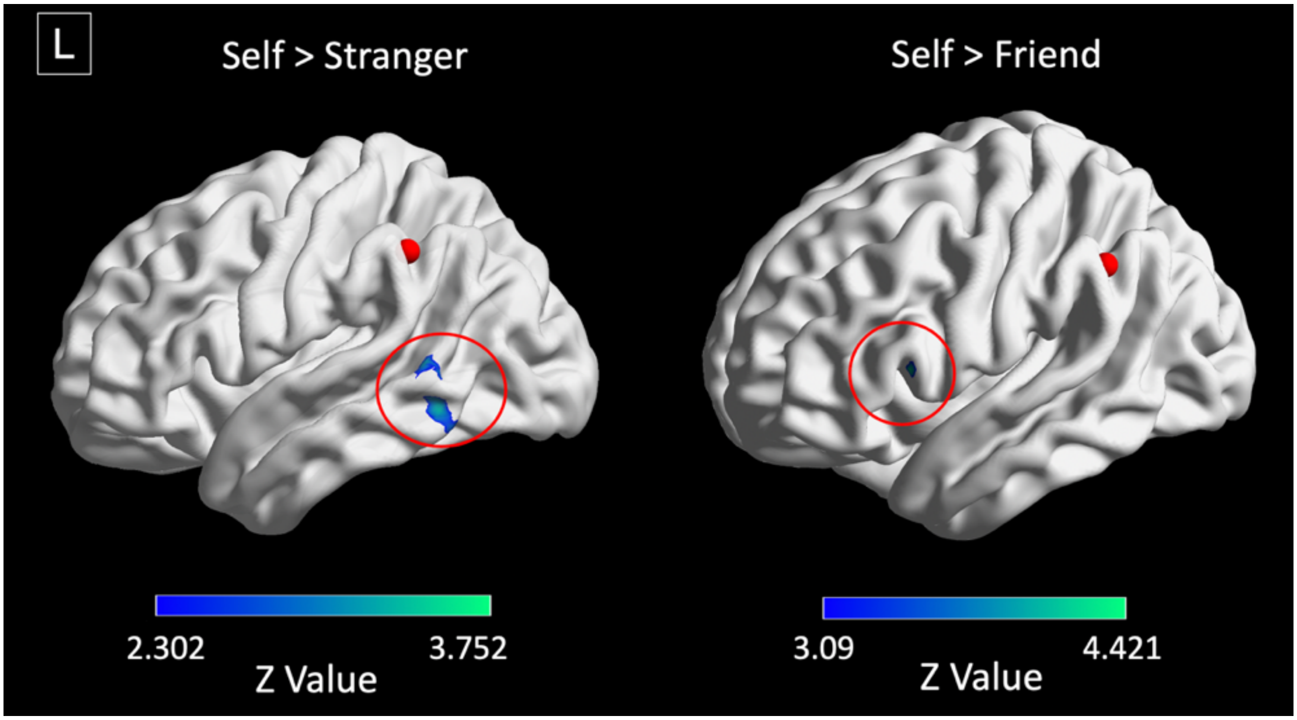


Fig 18. Task-modulated functional connectivity analysis of left IPL showed increased connectivity to temporooccipital regions during self-processing (self > stranger; cluster corrected $Z > 2.3$, $p < .05$) and the IFG (self > friend, $p < .001$, uncorrected).

gPPI Results: Functional connectivity of Left IPL with temporoccipital and frontoparietal regions

Contrast	Hemisphere	Area	*MNI (x,y,z)*	Max Z	Cluster Size	*P*
Self > Stranger	Left	Inferior Temporal Gyrus	(-54,-54,-10)	3.75	335	.00635
	Left	*Sub-maxima*: Inferior Lateral Occipital Cortex	(-52, -70, -14)			
	Left	*Sub-maxima*: Superior Temporal Sulcus	(-62,-48,8)			
	Left	*Sub-maxima*: Middle Temporal Gyrus	(-62, -50,4)			
	Left	*Sub-maxima*: Middle Temporal Gyrus	(-58, -52, -4)			
	Left	*Sub-maxima*: Posterior Middle Temporal Gyrus	(-60, -44, -2)			
Self > Friend	Left	Inferior Frontal Gyrus	(-54, 20, 10)	4.42	52	< .00001

Task-specific modulation during *self > stranger* also revealed increased connectivity along the insular and inferior frontal cortices. The gPPI analysis with the left insula seed (3mm radius sphere; centered at peak *x, y, z* = -44,12,6; from functional activation of *self >stranger*) was functionally connected to the left IFG pars opercularis (*x,y,z* = -52, 10, 16) with a small cluster ($n_{\text{vox}} = 8$) also in the right IFG pars opercularis (*x,y,z* = 54, 16, 14) ($z>2.3$, $p < .05$). Task modulation for *self > friend* at the more liberal threshold ($p < .001$ uncorrected) similarly revealed increased connectivity between the left insula and the Left IFG (left *x,y,z* = -54, 18, 10; right *x,y,z* = 58, 14, 10). Together, the connectivity results are indicative of recruitment of a wider neural circuitry for self-actions, extending from the temporooccipital regions to the frontoparietal regions.

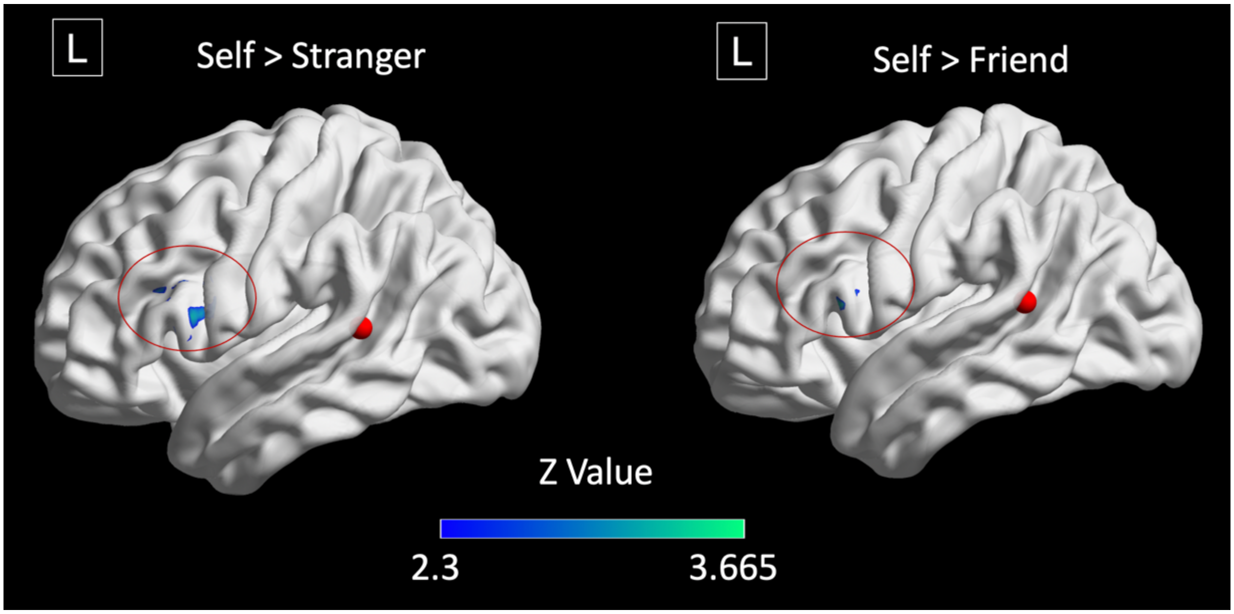


Fig 19. Task-modulated functional connectivity analysis of Left Insula showed increased connectivity to bilateral inferior frontal gyri (pars opercularis) (self > stranger; cluster corrected *Z* > 2.3, *p* < .05) and the Left inferior frontal gyrus (pars opercularis) (self > friend, *p* < .001, uncorrected).

gPPI results: Functional connectivity of the Left Insula and IFG

Contrast	Hemisphere	Area	*MNI (x,y,z)*	Max Z	Cluster Size	*P*
Self > Stranger	Left	IFG (pars opercularis)	(-52, 10, 16)	3.67	378	0.00156
	Right	IFG (pars opercularis)	(54, 16, 14)	3.35	240	0.0316
Self > Friend*	Left	IFG (pars opercularis)	(-54, 18, 10)	3.96	33	p < .001
	Right	IFG (pars opercularis)	(58, 14, 10)	3.38	8	p < .001

* Reduced threshold, *p* < .0001, uncorrected

ROI-MVPA

Using ROI-MVPA, we measured two-class decoding accuracy between each identity on the temporooccipital (localizer) and frontoparietal (meta-analytic) regions relative to chance performance (.50). We corrected for multiple comparisons using the false discovery rate (FDR) (Benjamini & Hochberg, 1995) on the number of ROIs (15 ROIs; $q < .05$, crit $p = .0260$). First, we measured the classification accuracy in the EBA derived from functional activity in the *bodies > chairs* localizer contrast. The right EBA showed significant above-chance decoding between all identities (self vs stranger, mean classification accuracy = .538, $t(19) = 2.816$, $p_{adj} = .025$, $d = .630$ self vs friend, mean classification accuracy = .535, $t(19) = 2.432$, $p_{adj} = .016$, $d = .688$, friend vs stranger, mean classification accuracy= .553, $t(19) = 3.115$ $p_{adj} = .016$, $d = .696$). The left EBA could also significantly distinguish between all identities: self and stranger identities, mean classification accuracy = .543, $t(19) = 3.993$, $p_{adj} = .0053$, $d = .893$, self and friend identities, mean classification accuracy = .535, $t(19) = 2.432$, $p_{adj} = .039$, $d = .544$ and friend and stranger, mean classification accuracy = .538, $t(19) = 2.691$, $p_{adj} = .030$, $d = .602$. The left pSTS also decoded between self and strangers, mean classification accuracy = .542, $t(19) = 2.639$, $p_{adj} = .031$, $d= .590$, but not between other identities (self vs friend, mean classification accuracy = .528, $t(19) = 1.746$, $p_{adj} = .121$, $d =.391$; friend vs stranger, mean classification accuracy = .524, $t(19) = 1.655$, $p_{adj} = .135$, $d =.370$). The right pSTS did not show significant decoding between any of the identities (*p*s > .11).

Since we had some difficulty extracting the right pSTS from the localizer for a few participants ($N = 4$), we also examined classification accuracy in the full probabilistic map of the atlas-derived right posterior superior temporal gyrus (pSTG) across all participants. We observed ~10% increase in training accuracy in the right pSTG and 12% increase in the left pSTG using the full probabilistic map. The testing accuracy also showed statistical significance using larger ROI regions. The right pSTG significantly decoded all identities (self vs stranger, mean classification accuracy = .558, $t(19) = 3.112$ $p_{adj} = .0159$, $d = .696$; self vs friend, mean classification accuracy = .554, $t(19) = 3.270$, $p_{adj} = .0159$, $d = .731$; friend vs stranger, mean classification accuracy = .546, $t(19) = 5.304$, $p_{adj} = .001$, $d = 1.186$). The left pSTG decoded the self from other identities (self vs stranger, mean classification accuracy = .567, $t(19) = 4.706$, $p_{\mathrm{adj}} = .0017$, $d = 1.052$; self v friend, mean classification accuracy = .571, $t(19) = 3.819$, $p_{adj} = .0074$, $d = .854$), but not friend vs stranger, mean classification accuracy = .525, $t(19) = 1.290$, $p_{adj} = .2332$).

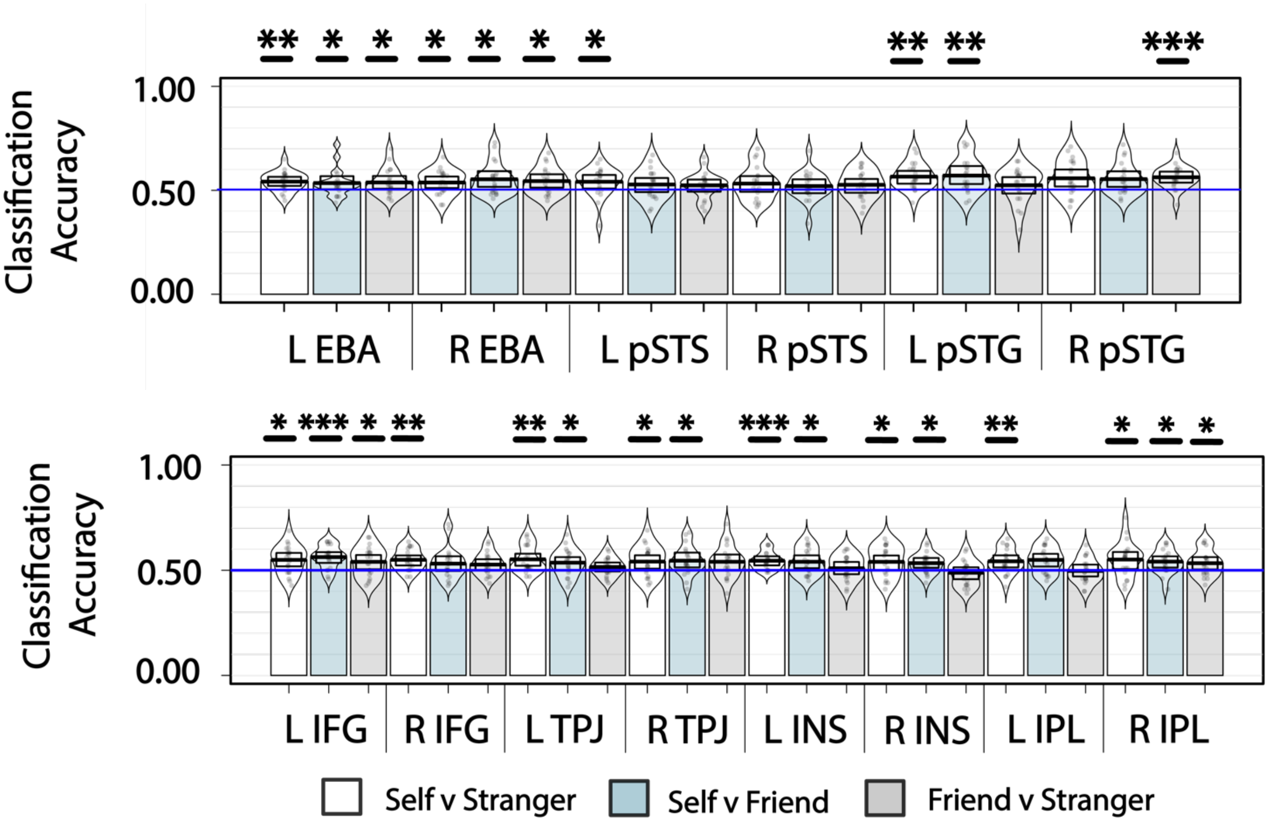


Fig 20. *Top:* Classification accuracies for temporooccipital regions of interest. *Bottom:* Classification accuracies for frontoparietal regions of interest. Average line denotes mean. White, light blue, and light gray fills indicate bar plots. Inference bands denote 95% Bayesian highest density interval with 1000 iterations. Horizontal blue line indicates chance-level decoding accuracy (.50). Abbreviations: extrastriate body Area (EBA); posterior superior temporal sulcus (pSTS); posterior superior temporal gyrus (pSTG); inferior frontal gyrus (IFG); temporoparietal junction (TPJ); insula (INS); inferior parietal lobe (IPL); * $p < .05$, ** $p \leq .01$, *** $p \leq .001$.

We then examined classification accuracy in the meta-analytic derived frontoparietal regions (*Figure 20)*. For ROI selection, we used the grid-search approach for selection of peak coordinate within a range of 4 voxels in 3D space. Across these ROIs, the left IPL significantly decoded self from other identities: left IPL in the parietal operculum/SMG (*x,y,z* = -50, -40, 26, self v friend, mean classification accuracy = .548, t(19) = 3.618, p_{adj} = .009, d = .809; self vs stranger, mean classification accuracy = .542, t(19) = 3.157, p_{adj} = .016, d = .809). The left IPL was attuned to the self and could not discriminate between the other identities (friend v stranger, mean classification accuracy = .496, t(19) = -.272, p_{adj} = .780). The right IPL (*x,y,z* = 52, -44, 32) significantly decoded all identities, self v friend, mean classification accuracy =.541, t(19) = 3.27, p_{adj} = .016, d = .731; self v stranger, mean classification accuracy = .551, t(19) = 2.683, p_{adj} = .031, d = .600); friend v stranger, mean classification accuracy =.532, t(19) = 2.414, p_{adj} = .039, d = .540). The bilateral TPJ as well showed selectivity for the self, *self* > *strange*r: (Left: *x,y,z* = -54, -32, 42, mean classification accuracy = .551, t(19) = 3.641, p_{adj} = .009, d = .814; Right: *x,y,z* =50, -24, 42, mean classification accuracy = .541, t(19) = 2.539 , p_{adj} = .036, d = .568), *self* > *friend*: (Left: mean classification accuracy = .536, t(19) = 2.518, p_{adj} = .036, d = .563; Right: mean classification accuracy = .546, t(19) = 2.961, p_{adj} = .019, d = .662), *friend* > *stranger:* (Left: mean classification accuracy = .516, t(19) = 1.582, p_{adj} = .150, d = .354; Right: mean classification accuracy = .540, t(19) = 2.295, p_{adj} = .116, d = .513). The bilateral insula further showed selectivity for the self. The left insula (*x, y, z* = -46, 0, 2) decoded the self: self vs stranger, mean classification accuracy = .545, t(19) = 4.995, p_{adj} = .001, d= 1.11, self vs friend, mean classification accuracy = .540, t(19) = 2.935, p_{adj} = .019, d= .656, but could not decode friend vs stranger, mean classification accuracy = .509, t(19) = .707, p_{adj} = .511. The right insula (*x, y, z* = 42, 6, 2) similarly decoded the self: self vs friend, mean classification accuracy = .534, t(19) = 3.146 , p_{adj} = .016, d

= .704 and self vs stranger, mean classification accuracy = .539, $t(19) = 2.409$, $p_{adj} = .039$, $d = .539$, but could not decode between friend vs stranger, mean classification accuracy = .487, $t(19) = -.901$, $p_{adj} = .406$.

As the meta-analytic coordinates localized the insula but did not include the inferior frontal cortices, we further measured identity discrimination in the IFG *pars opercularis* (IFG*po)* using the full probabilistic map generated from the Harvard-Cortical Atlas in FSL. We found significant discrimination for all identities in the left IFG*po,* self vs stranger: (mean classification accuracy = .549, t(19) = 3.211, $p_{adj} = .016$, $d = .718$), self vs friend: mean classification accuracy = .562, (t(19) = 5.169, $p_{adj} = .0012$, $d = 1.156$), friend vs stranger, mean classification accuracy = .540, $t(19) = 2.538$, $p_{adj} = .0361$, $d = .567$. We found significant decoding for self vs stranger in the right IFG*po*, mean classification accuracy =.549, $t(19) = 4.446$, $p_{adj} = .0025$, $d = .994$, but not self vs friend, mean classification accuracy = .531, $t(19) = 1.859$, $p_{adj} = .1072$, or friend vs stranger, mean classification accuracy = .527, $t(19) = 2.260$, $p_{adj} = .0519$. The bilateral TPJ/SMG (left *x,y,z*: -54, -32, 42; right *x,y,z*: 50, -24, 42) could also discriminate self vs strangers (left: mean classification accuracy = .551, $t(19) = 3.641$, $p_{adj} = .009$, $d = .814$; right: mean classification accuracy = .541, $t(19) = 2.539$, $p_{adj} = .036$, $d = .568$) and self v friends (left: mean classification accuracy = .536, t(19) = 2.518, $p_{adj} = .036$, $d = .563$; right: mean classification accuracy = .546, $t(19) = 2.961$, $p = .019$, $d = .662$). Friend vs strangers were not significantly decoded from each other (left: mean classification accuracy =.516, t(19) = 1.582, $p_{adj} = .150$; right: mean classification accuracy = .540, $t(19) = 1.79$, $p = .116$).

As a control for predictive power, we used the grid-search approach in the cerebellum centered on (*x,y, z* = -18, -70, -54) and did not identify significant identity decoding between our identities of interest (both self vs friend and self vs stranger). Out of 60 possible locations, one coordinate (*x,y,z* = -20, -68, -46) did show initial significance for both identity decodings, though it disappeared after the FDR correction (self vs stranger: p_{adj} = .059; self vs friend: p_{adj} = .037).

Representational Dissimilarity Analysis (RDA)

Based on the significant identity decoding revealed in the temporooccipital and frontoparietal regions from MVPA analysis, we implemented searchlight RDA analyses to explore the featural space of low and mid-level signatures to recognition of actions performed by self and others. Multiple RDA searchlight analyses were conducted using body structure based on limb lengths, movement distinctiveness (DTW), speed, and acceleration differences as behavioral RDMs, as well as theoretical RDMs based on expected visuomotor familiairty. No relationships were found between speed or acceleration differences to neural similarity, however significant relationships were found to both DTW and body structure, as well as the theoretical RDMs.

DTW RDA: Movement Distinctiveness

As shown in Figure 21, the whole-brain RDA searchlight between the DTW RDM and neural RDM produced significant bilateral activity across the action observation network. Large clusters were found in the early visual cortex, spanning the early visual cortex (V1) to the temporooccipital regions including the fusiform gyri and EBA (*x, y, z* = 24, -82, -22). Clusters were also found in the bilateral anterior supramarginal gyrus and the IFC (spanning pars opercularis to the insular cortex), as well as in the bilateral premotor cortex (*x,y,z* = 18, -10, 52) and dorsolateral prefrontal cortex.

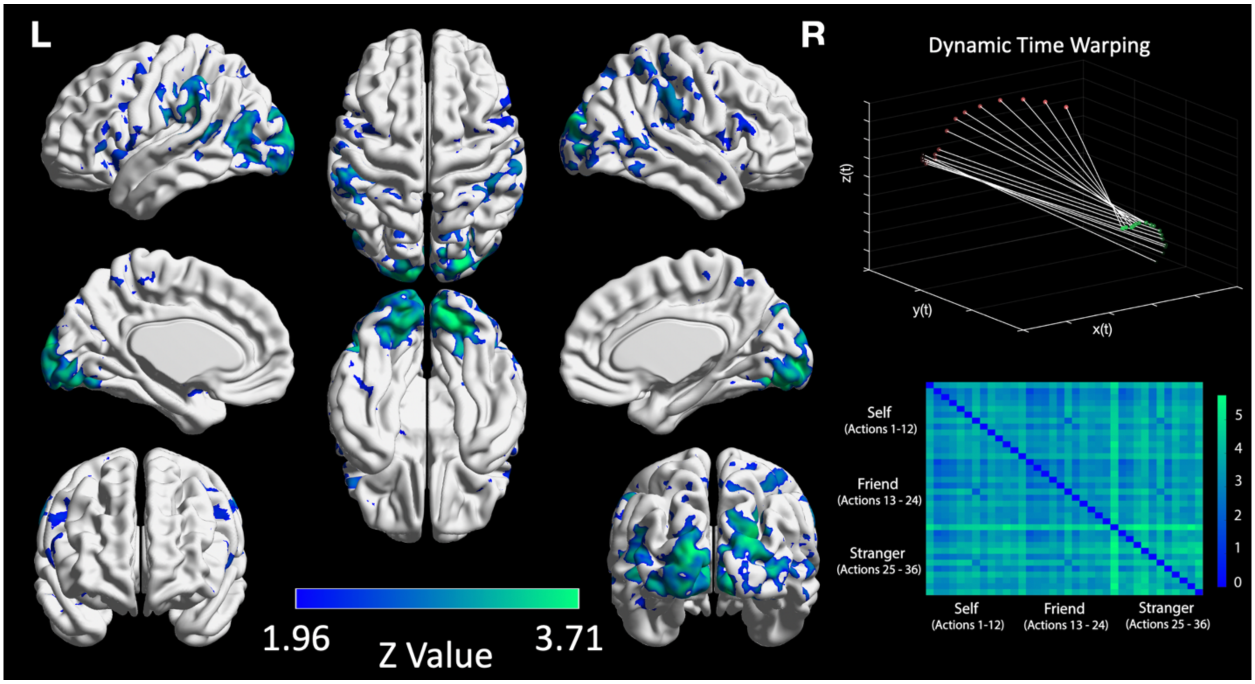


Fig 21. Searchlight RDA results depicting the z-transformed activity map for significant correlations between the DTW behavioral RDM and the neural RDM based on activity patterns for actions across all identities (self, friend, stranger). Activation map reflects TFCE-corrected brain activity after 10000 non-parametric Monte Carlo simulations. *Top Right*: DTW figure showing one joint from one actor's action time series (in red) with lines measuring similarity to the corresponding joint in another actor's time series to find the optimal decrease in dissimilarity over time. *Bottom Right:* Behavioral RDM used for DTW RDA reflecting the Euclidean distance between identity/action categories. Darker blue indicates smaller distance (i.e., more similarity).

The results of the searchlight RDA with DTW were largely restricted to temporooccipital regions, though some neural activity was elicited more anteriorly in the IPL and IFC. Hence, we computed the degree to which univariate task average for the self in the left-lateralized IPL and IFC/Insula overlapped with the feature space of the DTW RDA map. Movement dissimilarity could explain portions of task-evoked activity in the left IPL (58% voxel overlap for *self* > *stranger*; 49% voxel overlap for *self* > *friend*), with the overlap localized to the anterior portions of the SMG. The remaining area of posterior voxels of the Left SMG did not overlap. For the univariate task-evoked left IFC we found 62.5% overlap with the task-evoked *self* > *stranger* neural activity. However, given the relatively small cluster size of the task-evoked IFC, we used the full probabilistic ROI for the IFG that showed significant decoding of the self from other identities, which revealed much lower overlap (36% overlap in the left IFG and 33% in the right IFG).

For the temporooccipital regions, we measured the percent overlap for each participants' multivariate localizer ROI (normalized to standard space) and the group-level DTW RDA map. We found significant overlap of the EBA temporooccipital localizers with the DTW RDA map, consisting of 60.5% of the Left EBA localizer, 45% of the Right EBA localizer, as well as overlap in the pSTS: 28.7% of the Left pSTS localizer, 13.5% of Right pSTS localizer. As the probabilistic map of the pSTG showed significant identity decoding at the multivariate level, we further measured its overlap with the DTW RDA neural patterns and found 27.7% of voxel overlap of the left pSTG and 13.6% of the right pSTG with the RDA map. Together, the results suggest the contribution of movement distinctiveness as a factor to self-action recognition across regions of the AON, though not accounting for the entirety of self-processing.

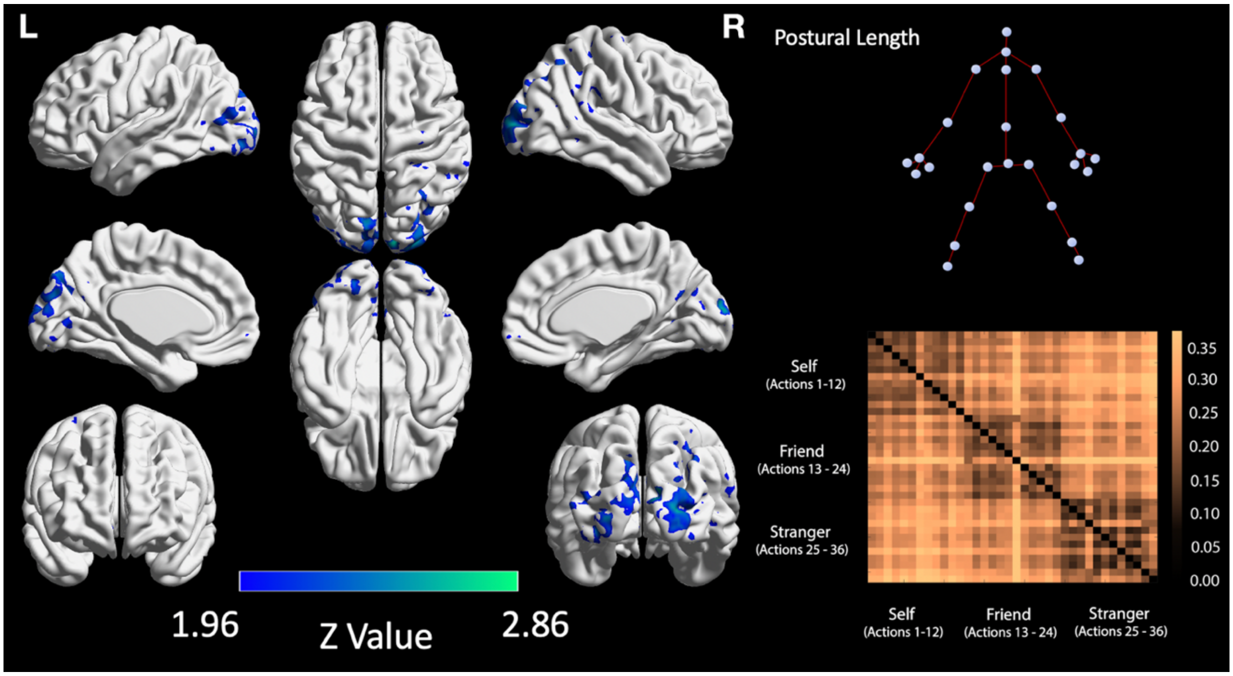


Fig 22. Searchlight RDA results depicting the z-transformed activity map for significant correlations between the body structure (postural length) behavioral RDM and the neural RDM based on activity patterns for actions across all identities (self, friend, stranger). Activation map reflects TFCE-corrected brain activity after 10000 non-parametric Monte Carlo simulations. *Top Right*: Body structure showing point-light joints interconnected to comprise 24 limbs (red lines). Pairwise distances for each limb length were calculated between each action for each identity. *Bottom Right:* Behavioral RDM used for body structure RDA reflecting the Euclidean distance between identity/action categories. Darker colors indicate smaller distance (i.e., more similarity).

Body Structure RDA: Postural Length

For the contribution of body structure to the feature space of identity recognition, we conducted RDA across the whole-brain using postural length as the target RDM. We identified activity patterns largely in the occipital and temporal regions, localized to the early visual regions extending to lateral occipital regions. Except for the bilateral occipital (i.e., early visual) regions, we noticed patterns of activity present at higher-level regions that were more right-lateralized, including small activity patterns along the superior parietal lobe to the right postcentral gyrus to the superior frontal gyrus, and a small cluster in the medial prefrontal cortex. We then calculated the percent activity overlap with ROIs from the AON. We found no overlap in the task-evoked left

IFG or IPL for the self, no activity overlap in the localizer for the left pSTS, and negligible overlap in the right pSTS (0.89%). We did, however, identify patterns of activity in the left EBA (18.1%) and the right EBA (11.24%) that represented body structure across identities.

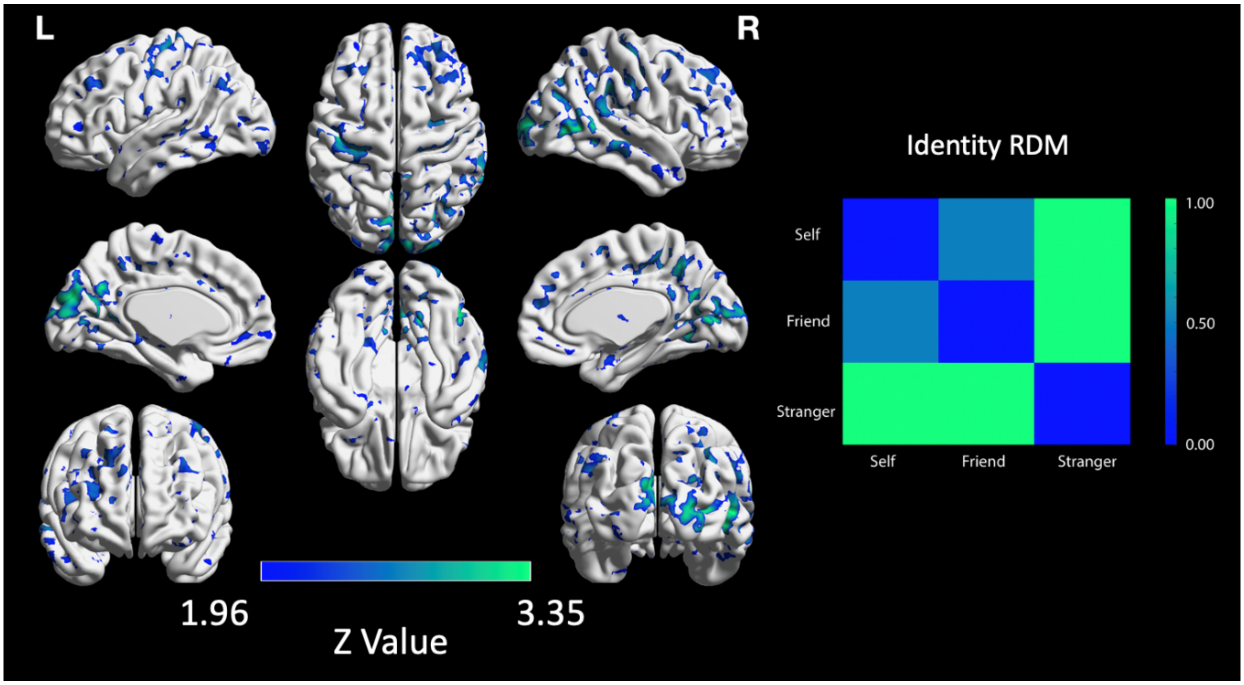


Fig 23. Searchlight RDA results depicting the z-transformed activity map for significant relationships between the theoretical identity RDM (self: 0, friend: .5, stranger: 1) and the neural RDM based on activity patterns for actions across all identities (self, friend, stranger). Activation map reflects TFCE-corrected brain activity after 10000 non-parametric Monte Carlo simulations. *Right:* Theoretical RDM used for the RDA reflects the Euclidean distance between identity/action categories. Darker blue indicates smaller distance (i.e., more similarity).

Identity RDA: Featural Overlap

Finally, we included a theoretical RDM based on the identity matrix computing visuomotor familiarity to oneself across the whole-brain (*Figure 23*). Large clusters were present in the bilateral postcentral gyri extending to the precentral gyri and supplementary motor areas. We also found clusters in the bilateral early visual cortices extending through the temporooccipital cortex (including EBA, MTG). Other clusters were found in the precuneus, right dorsolateral prefrontal cortex, left insula (direct overlap with the meta-analytic coordinate for the left insula: x,y,z = -46, 0, 2), bilateral frontal operculum, and bilateral IPL. Considering the large number of AON regions that showed significant activity patterns from the identity RDA, we used the task contrast generated from FSL's RFT cluster correction for *self > stranger* and measured self-related overlap. We found greatest overlap in the right ventrolateral prefrontal cortex (42%), right dorsolateral prefrontal cortex (36%), right IPL (36%), and right pSTS (25%).

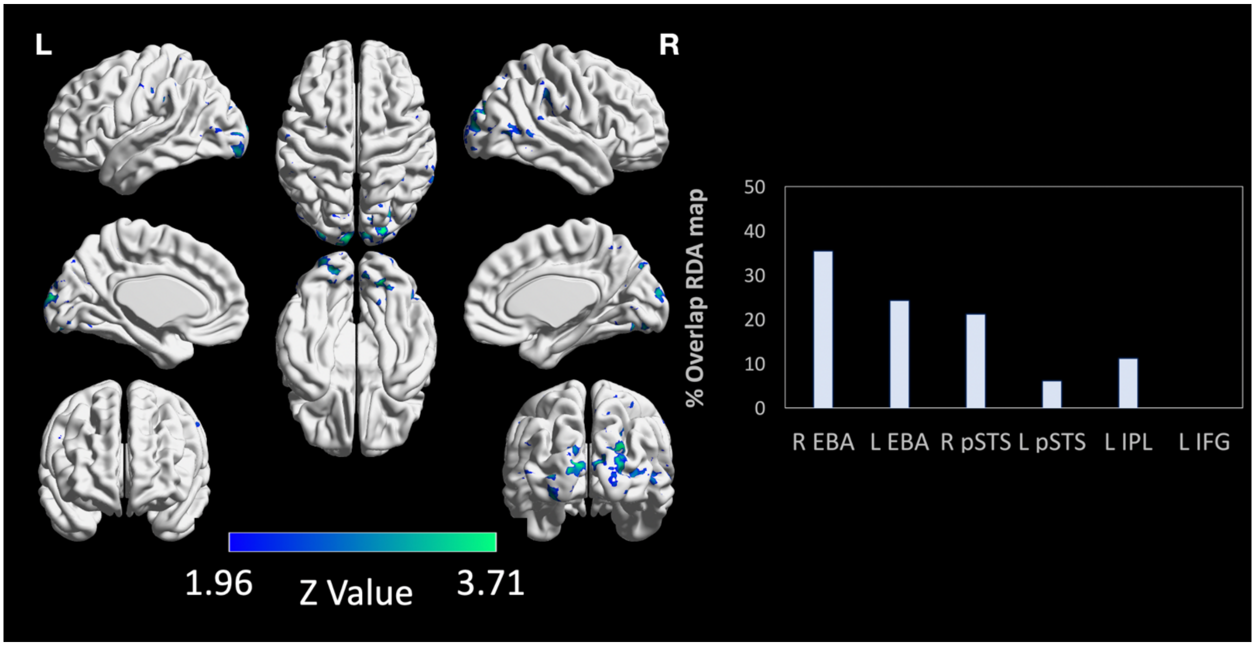


Fig 24. Searchlight RDA results depicting the z-transformed activity map for significant relationships between the intersection of activity maps for the theoretical identity RDA and the movement distinctiveness DTW RDA across all identities (self, friend, stranger). Activation map reflects TFCE-corrected brain activity after 10000 non-parametric Monte Carlo simulations. *Right:* Graph shows percent overlap in ROIs that were either task evoked in the frontoparietal regions (Left IPL; L IFG) or generated from the localizer in the temporooccipital regions (pSTS; EBA). The DTW and identity overlap shows a largely parametric trend, with little overlap in the left-lateralized frontoparietal regions, and large increases in the bilateral temporooccipital regions.

We then separately computed the intersection of activity maps between movement distinctiveness and body structure with the identity matrix RDA map. For the intersection between DTW (movement distinctiveness) and identity RDA maps (*Fig 24*), we found overlap largely localized to the bilateral occipital regions spanning the lateral occipital cortices and temporal regions (middle temporal gyrus). A few small clusters of activity were also found in the bilateral middle frontal gyrus, parietal operculum (spreading up to the posterior SMG), posterior cingulate cortex, right precuneus and superior parietal lobule. Small patterns of activity were also found in the more anterior part of the IPL (bilateral anterior SMG), lateralized slightly more to the right. Notably, we did not find overlapping activity patterns between the task-evoked left IFG activity for the self and the intersection map between DTW and identity, and only little overlap with the task-evoked left IPL, which was localized more anteriorly (*self > stranger:* 11.4%, *self > friend:* 10.8% overlap), despite both the task-evoked Left IFG (62.5%) and Left IPL (58.4%) eliciting significant overlap with the DTW RDA. We also found little overlap with the left pSTS localizer (6%). However, we did find considerable overlap between the EBA localizer and the intersection map between DTW RDA and identity RDA (right: 39%; left: 24%), as well as the right pSTS (21.7%). These results together suggest that regions of the bilateral EBA and right pSTS rely on movement distinctiveness as a cue to identity. However, the frontoparietal regions, though similarly attuned to action features as shown by the neural relationship to movement distinctiveness, did not appear to predominantly rely on these action features as cues to identity.

Finally, the intersection map produced between the RDAs for posture and identity were largely localized to the EBA. We found that the majority of voxels that represented posture in the right EBA overlapped with identity (62.25%), and a smaller, though considerable, percentage of voxels representing posture in the Left EBA overlapped with identity (45%).

Discussion

Our findings converge on a specialized neural circuitry for visual self-recognition for actions. This circuitry was situated across the action observation network (AON) involving the temporooccipital regions and frontoparietal regions. Four main findings are discussed below: (1) AON was recruited across all identities (2) temporooccipital regions within the AON were involved in domain-general identity decoding (3) frontoparietal regions within the AON were broadly attuned to self-recognition (4) visual self-action recognition additionally recruited motor planning regions.

First, the behavioral results confirmed that all identities could be discriminated above chance performance in PLDs. Self-recognition performance also showed no relationship at the behavioral level to movement distinctiveness or with any of the other lower-level visual features (speed, acceleration). We then moved to the univariate analyses where we identified common recruitment of regions of the AON regardless of identity. This is indicative, at the univariate level, of similar neural systems at play for actions of self, friends, and strangers. These regions spanned the inferior frontal gyrus (IFG), inferior parietal lobe (IPL), and supplementary motor area (SMA)— traditionally characterized as part of a larger mirror neuron system (e.g., Rizzolatti & Craighero, 2004; Iacoboni, 2009; Koski et al., 2003; Mukamel et al., 2010). This corroborates a long line of empirical work explicating the common visuomotor codes that underlie the self and others (e.g., Casile & Giese, 2006; Rizzolatti & Craighero, 2004; Iacoboni, 2009; Iacoboni & Dapretto, 2006). Importantly, these theories often suggest actions are a form of "sensemaking"— that it is through the motor experience of performing movements that we are able to grasp and embody the meaning of actions of others (e.g., Rizzolatti and Craighero, 2004). In our task that involved observing our own actions, the privileged familiarity of motor experience in performing the action may be the defining factor that enhances and allows for self-other differentiation in visual recognition.

Yielding support for this, we identified a primarily left-lateralized frontoparietal network for self-recognition, shown both by univariate increases in task-specific activity, as well as task-modulated increases in functional connectivity within the AON. Our activity is notably more left-lateralized than earlier work on self-recognition that have used static body images, faces, or isolated dynamic body parts, that typically show a degree of right-hemisphere lateralization (Brady et al., 2004, Keenan et al., 2003; Platek et al., 2004; Platek et al., 2003; Hodzic et al., 2008; Vocks et al., 2010; Uddin et al., 2006; Macuga and Frey, 2011; cf Uddin et al., 2005; cf Turk et al., 2002; cf Sperry et al., 1979). Our results introduce the possibility that processing of dynamic stimuli related to our whole-body may undergo more left-lateralized routes to awareness. This falls in line with laterality differences depicted in clinical conditions that have damage to the left hemisphere, such as in limb apraxia (e.g., Haaland et al., 2000; Haaland, 2006; Watson et al., 2019), as well as in empirical work using dynamic bodily stimuli of the self that similarly show a pattern of left-lateralization for self-recognition (e.g., Bischoff et al., 2012; Wozniak et al., 2021). However, it is also important to note that while our study did not yield bilateral involvement of the IPL or bilateral involvement of other AON regions when using the non-parametric univariate thresholding, we did find bilateral involvement of the IPL and the AON regions when using FSL's RFT univariate cluster correction (reported in Appendix) as well as in our multivariate analyses. At the univariate level, we chose to interpret the more conservative non-parametric results (TFCE), as TFCE been shown to more sensitive yet less prone to false positives in the literature (Smith & Nichols, 2009), which resultingly eliminated the activity in the right hemisphere. Thus, the left-laterality from our results cannot be stated with certainty, but does however indicate that the network for self-action recognition is not primarily right-lateralized as previously suggested in some self-recognition work.

The two key regions we found, the left IPL and IFG/inferior frontal cortex (IFC), are not only causally implicated in conditions linked to motoric planning or production (e.g., limb apraxia), but further comprise core nodes of the mirror neuron network, containing specialized visuomotor neurons that map the perception of others' actions to our own motor system. Studies on visual self-recognition that have similarly shown recruitment of the IPL and IFC often use familiar stimuli including faces (Uddin et al., 2005; Uddin et al., 2006) or voices (Kaplan et al., 2008). Task recruitment of these regions is largely attributed to greater mirror neuron activity for embodied self-representations based on a common coding mechanism that matches common perceptual and motor codes (Uddin et al., 2007; Uddin et al., 2005; Uddin et al., 2013; Prinz, 1997). Since self-generated actions would presumably have the greatest visuomotor match due to privileged motor experience performing the action, these results fall in line with the above common coding theory (Prinz, 1997; Hommel, 2001) and underscore the importance of the motor system to visual self-recognition.

While these mirroring regions were involved for processing all identities, they notably showed greatest specificity for the self. We found a central role of the left IPL for self-processing, possibly acting as a hub/integratory zone—revealed both by increases in univariate activity and increased functional connections to both temporooccipital regions and bilateral inferior frontal cortices during self-processing. Connectivity between the more posterior portion of the supramarginal gyrus (SMG) of the left IPL increased in task-modulated functional connectivity during self-processing with the left-lateralized temporooccipital regions, including the inferior temporal gyrus, lateral occipital cortex, and superior temporal sulci. The pattern of results falls in line with recent characterizations that underscore the possibility of recurrent patterns of connectivity between the left posterior parietal cortex and temporooccipital regions (left EBA) (Bracci et al., 2015) that facilitate a matching mechanism (individually defined as comparator regions) computing a discrepancy signature between “seen” and “felt” bodily representations (Limanowski & Blankenburg, 2017). Specifically, it is thought that the EBA and posterior parietal cortex together form a specialized pathway attuned to the visuoproprioceptive congruency of information- for example, when the observed hand orientation matches the participants’ hand orientation, task-modulated functional connectivity between these regions should increase due to the increase in congruency. Hence, increases in functional connectivity during visuoproprioceptive congruencies could facilitate the more advanced positional and bodily estimation computations required for self-recognition, which naturally have the greatest visuoproprioceptive congruency (Limanowski & Blankenburg, 2017). In line with this, specialized network characterizations increasingly implicated in the processing of self-related static stimuli such as our own faces or bodies, have also been shown to recruit specific cortical areas that are different from domain-general processing of the stimulus (Platek et al., 2006; Sugiura et al., 2005; Hodzic et al., 2009), such as regions unique for self-face recognition rather than for identity-general face recognition. Since the prior work largely uses isolated faces, body parts, or static stimuli, our study involving holistic movements of the body further extends the circuitry to more widespread recruitment of the AON spanning the temporooccipital cortex to the IFC. Specifically, task-modulated activity for the self was functionally connected from the left temporooccipital regions to the IPL, and from the left IPL to the left IFC (spanning the insula)—indicative of a broader neural circuitry across the AON for the self-recognition of the dynamic whole-body.

Across methodologies, our findings converge on a key role of the posterior IPL to visual self-recognition. The IPL is a large cortical region consisting of posterior regions involved in person identity decoding (Tholen et al., 2019), construction of integratory body schemas (Jackson and Decety, 2004), reference frame transformation (Gallivan & Culham, 20015), body related processing (Ehrsson et al., 2005, Pellijeff et al., 2006, Shimada et al., 2005), modality-independent action processing (Wurm & Lignnau, 2015; Urgen et al 2019), action-outcome monitoring (Van Kemenade et al., 2017; Van Kemenade et al., 2019), self-processing (e.g., Bréchet et al., 2018, Uhlmann et al., 2019; Uddin et al., 2005; Uddin et al., 2007), and is spatially expanded in primates (Orban et al., 2004). Urgen et al (2019) delineated the central role of the IPL involved in action representations—building upon the visual representations of actions and transforming them into more abstract and semantic representations coding action type and intention. As person-identity labeling is abstracted relative to featural processing of actions, the IPL involvement for self-actions maintains contiguity with the logic. Our activity for the self in the IPL was localized more posteriorly in the left supramarginal gyrus, including parts of the angular gyrus, with direct overlap with the mirror neuron system (supramarginal gyrus in IPL). As follows with the more posterior activation of the parietal cortex, this establishes a close link to studies that have shown the importance of the posterior parietal cortex in motor planning, as well as in generating the forward models used in motor control theories as discrepancy comparators to predict the outcome of our own actions (Desmurget et al., 1999). Motor planning is hence privileged only to oneself, and regions that are involved in motor planning are by consequence attuned to the self, thus indicative of a key role of the IPL for attributing self-generated actions to oneself based on motor familiarity.

As the IPL is a functionally heterogenous region, some degree of perceptual dependence has been shown, though localized more anteriorly (e.g., Goldenberg & Spatt, 2009; Wurm &

Caramazza, 2021). Results from the representational dissimilarity analysis (RDA) that measured the relationship between movement distinctiveness of actions and neural activity patterns across the whole-brain confirmed some of this dependence, as more anterior portions of the IPL represented features related to movement distinctiveness as well as did regions in the temporooccipital cortex. Moreover, posterior portions of the IPL overlapped with the activity patterns derived from the visuomotor identity RDA, indicative of higher-level identity information represented in the activity patterns, rather than on lower-level visual action features. In keeping with prior work, posterior portions of the IPL are shown to be involved in conceptual processing and higher-level action and conceptual knowledge (Leshinskaya & Caramazza, 2014; Wurm & Caramazza, 2021) as well as bodily self-awareness (i.e., in the angular gyrus; Bréchet et al., 2018). Thus, the pattern of results confirms the existence of functional subregions within the IPL that may be attuned to both visuomotor features of the action identities (anteriorly) as well as conceptual-level identity decoding specialized at some level for the self (posteriorly).

In a recent study, the involvement of motor experience was eliminated from the research design (Wozniak et al., 2021). Here, the researchers arbitrarily associated an identity (self, friend, or stranger) to point-light actions of one actor as a means of eliminating not only visual familiarity, but also motor/proprioceptive familiarity. The researchers identified a similar set of left-lateralized regions: IFG pars opercularis, middle frontal gyrus (MFG), and SMA. Notably however, they did not find recruitment of the IPL. In relation to our study, which similarly reduced visual cues to identity as in PLDs, but still maintained the natural proprioceptive familiarity we have with our bodies, the presence of IPL (posterior SMG) activity and its decoding ability in our study further underscores the centrality of this region to enacting motor codes of the self to visual self-recognition.

Across the multivariate regions of interest, the majority of temporooccipital regions (bilateral EBA, right pSTG) could discriminate between all identities while the frontoparietal (IFG, IPL, TPJ) regions were largely attuned to decoding the self. Note however, the pattern did not maintain for three regions: left IFG, right IPL, and left pSTG— whereas the left pSTG was specialized for the self (possibly due to its role in reafference: Iacoboni et al., 2001), the left IFG and right IPL could decode all identities. Holistically, however, the results suggest that the majority of the temporooccipital regions could be involved in initial identity decoding based on low and mid-level bodily action features, while the frontoparietal regions likely involve distinct routes for processing of self-actions at the motoric/proprioceptive level. For the temporooccipital regions, the literature has shown that the bilateral EBA is implicated in general identity processing related to perceivable features, such as in morphological cues from body images (De Bellis et al., 2017; Hodzic et al., 2009), where disruptions to the EBA significantly impair general body identity discrimination (Urgesi et al., 2007), reflective of the importance of EBA to person-identity processing, rather than a specific preference for the self (e.g, Hodzic et al., 2009; Sasaki et al., 2018; cf Jeannerod, 2004; cf David et al., 2007).

How do we account for the possibility that self-generated actions are variable and could convey more idiosyncratic visual features that allow for self-recognition? As posture and speed provide critical cues to valanced processing, such as the perception of threat (Candidi et al., 2011) or emotion (e.g., De Gelder et al., 2006; 2009), it is similarly possible that these idiosyncratic and "valenced" features related to oneself— distinctive visual cues— may pop out in self-awareness. Using RDA, we explored the featural space represented in neural activity patterns based on the contribution of low- and mid-level action features to self-recognition. We found no relationships between neural activity patterns across the whole brain with speed or acceleration. However, movement distinctiveness showed clear relationships to many regions within the AON. We identified a primary role of the bilateral occipital and temporal regions in relation to the spatiotemporal movement distinctiveness between actions.

Clusters of other regions throughout the AON, including in the IPL and IFG further related to movement distinctiveness of the actions, though with non-overlapping portions of these regions specialized for self-processing that were not based on low or mid-level features related to movement distinctiveness, speed, acceleration, or posture. Hence, visual variability in relation to the actions could serve as an early cue to self-identity but cannot account for the totality of activity in the AON, as numerous voxels of the supramarginal gyrus of the IPL and IFG that emerged from task-based self-processing did not overlap with neural activity patterns of movement distinctiveness. Many of these non-overlapping voxels in the IPL, however, did overlap with the identity processing, measured in the RDA analyses. Furthermore, task-evoked self-processing appeared to recruit activity in the posterior portions of the SMG nearer to the angular gyrus, while distinctiveness related to action features appeared to "spread" more anteriorly in the SMG. Note that the overlap present in the RDA does not necessarily suppose that the same populations of neurons performing identical functions (e.g., Peelen and Downing, 2005), as the analysis stands largely as a correlational measure. In a broader sense, our results specify that these overlapped regions from the RDA likely represent interleaved populations of neurons encoding functionally similar properties.

We did not find much overlap with the left IFG (pars opercularis) and activity patterns from the identity RDA (though sparse overlap was present in pars triangularis), while large portions of these regions of the IFG were accounted for by the DTW RDA. Hence, while the RDA analyses revealed activity patterns in the IPL and IFG that represented movement distinctiveness, the RDA showed that the representational space for identity processing in these regions did not overlap with movement distinctiveness or posture. Hence while the left IFG may be attuned to stimulus features related to movement distinctiveness, the patterns of activity reflected in the representational space of the region did not overlap with identity processing, indicative of other factors beyond action features driving identity decoding in the IFG. These other factors could like likely be top-down conceptual contributors to identity representations (Sokolov et al., 2018) in this region that instead pertain more to evaluative judgments of the self (Kaplan et al., 2008; Devue et al., 2007). It is important, however, to reiterate the functional heterogeneity of these regions (e.g., Weiner & Grill-Spector, 2011; Graziano et al., 2002; Graziano, 2016). Deeper levels of decoding at the level of each ROI are necessary to understand their exact processing pathways and connectivity with other regions, using both ROI representational metrics, as well as directed measures of connectivity between these subregions, which are current limitations in our study.

The bilateral EBA revealed significant overlap with both movement distinctiveness and identity, suggesting the movement distinctiveness serves as a driving cue to identity in the EBA, which could then be transformed and passed upward to interconnected areas in the frontoparietal regions (Ma et al., 2018). Interestingly, the identity-based RDA also produced large clusters in motor planning and execution regions including M1, SMA, S1, the premotor cortex, as well as the ventrolateral prefrontal cortex, which is known for its involvement in motor inhibition (e.g., Leung and Cai, 2007; Aron et al., 2003). This overlap suggests that person identity representations based on visuomotor familiarity are coded in these motor planning and motor inhibition regions. Considering the relationship between neural activity patterns in the ventrolateral prefrontal cortex and identity processing, one conjecture could be that the region functions in isolation or in conjunction with larger circuitry (e.g., including the dorsolateral prefrontal cortex) to inhibit automatic/imitative motor responses automatically evoked when viewing our own and actions of others (e.g., Brass, Zysset, Von Cramon, 2001; Darda & Ramsey, 2019).

The temporooccipital regions were also functionally connected to the IPL during task-modulated self-processing, again indicative of the central role of the left IPL as a hub for self-identity. Information flow between temporooccipital regions and the IPL could provide a pathway to assist in the process of self-other differentiation prior to higher-level self-other differentiation in more posterior parts of the IPL (Limanowski & Blankenburg, 2017). Further, without the control for visual familiarity (i.e., measuring the self > stranger task contrast), we found enhanced connectivity between the left IPL and the temporooccipital regions including a smaller cluster in the anterior STS/MTG and lateral occipital cortex (EBA). Collectively, based on the RDA and connectivity analyses, it appears that course identity decoding based primarily on visual properties of the stimulus could occur in the temporooccipital regions and fed upward for integration in the IPL. Identity information from the IPL could then further flow upward through feedforward connections to the IFG (Sasaki et al., 2018), and/or influenced by backward connections from the IFG, evidenced by the task-modulated functional connections between the IPL seed region and the bilateral inferior frontal cortices. As the left IFG is also shown to be involved in more evaluative and autobiographical tasks (Morin and Michaud, 2007) as well as independent of the stimulus domain, it could plausibly enact a top-down and evaluative role (Soch et al., 2017) for identity processing, potentially explaining why the IFG (left) could decode all identities at the multivariate level. However, given the non-directionality of our connectivity analysis, our interpretation is largely speculative in determining directionality of information flow.

For discussion purposes, a cluster of activity in the ACC was also observed during the univariate task contrast of self > stranger actions. We did not extensively analyze the involvement of the ACC, though many studies underscore its importance to self-awareness and involvement in surprise signaling—as a sort of prediction error that emerges when viewing oneself (Apps & Tsakiris, 2014). The ACC is also a key node in the salience network which may explain why univariate activity spanned the insula during self-processing. Notably, however, we did not find increases in functional connectivity between the insula and ACC during self-processing, pointing less to salience network recruitment, and more to evinced activation along the inferior frontal cortices shown by the increased functional connectivity between the insula and IFC during self-processing. As the insula has a well-known role in bodily and interoceptive states of the self (anteriorly), the pattern of results could also suggest more intrinsic bodily attunement during self-recognition, or even more reflective processes related to the self, considering its proximity and anatomical connections to the ventromedial prefrontal cortex, a region well-characterized to code for social and reflective processes of oneself (e.g., Lieberman et al., 2019).

Human beings are embedded in dynamic contexts that change across time and space. The contribution of motion information from the body shapes the ability to recognize not only the actions of others (Yovel & O'Toole, 2016), but also our own. Here, we characterize a cortical network across the AON that prioritizes for self- actions. These findings together reveal how awareness of our own actions taps into more than pure visual familiarity. Self-recognition shares common sensorimotor links to others, as well as specialized processing attuned to individual visuomotor contributions.

Acknowledgements

We thank Sophia Baia and Kelly Xue for assistance with data collection and stimuli creation, and Elinor Yeo, Jolie Wu, Kelly Nola, Nicolas Jeong, Danya Elghebagy, David Lipkin, Shahan McGahee for assistance with stimuli creation. We thank Jeff Chiang for helpful advice on the analyses. This project was supported by GRANT to H.L., GRANT to M.M, and APA Dissertation Award to AK. A preliminary version of this project was presented at the Virtual Society for Neuroscience (2020) and V-Vision Sciences Society (2020).

References

Apps, Matthew AJ, and Manos Tsakiris. "The free-energy self: a predictive coding account of self-recognition." Neuroscience & Biobehavioral Reviews 41 (2014): 85-97.

Aron AR, Fletcher PC, Bullmore ET, Sahakian BJ, Robbins TW (2003) Stopsignal inhibition disrupted by damage to right inferior frontal gyrus in humans. Nat Neurosci 6:115–116.

Astafiev, S. V., Stanley, C. M., Shulman, G. L., & Corbetta, M. (2004). Extrastriate body area in human occipital cortex responds to the performance of motor actions. Nature neuroscience, 7(5), 542-548.

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. Perception, 33(6), 717-746.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. [erratum appears in J Autism Dev Disord 2001 Dec;31(6):603]. Journal of Autism & Developmental Disorders.

Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2003). General multilevel linear modeling for group analysis in FMRI. Neuroimage, 20(2), 1052-1063.

Beardsworth, T., & Buckner, T. (1981). The ability to recognize oneself from a video recording of one's movements without seeing one's body. Bulletin of the Psychonomic Society, 18(1), 19-22.

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. Annals of statistics, 1165-1188.

Blanke, O. (2012). Multisensory brain mechanisms of bodily self-consciousness. Nature Reviews Neuroscience, 13(8), 556-571.

Blanke, O., Slater, M., & Serino, A. (2015). Behavioral, neural, and computational principles of bodily self-consciousness. Neuron, 88(1), 145-166.

Bischoff, M., Zentgraf, K., Lorey, B., Pilgramm, S., Balser, N., Baumgartner, E., ... & Munzert, J. (2012). Motor familiarity: Brain activation when watching kinematic displays of one's own movements. Neuropsychologia, 50(8), 2085-2092.

Bracci S, Caramazza A, Peelen MV (2015) Representational similarity of body parts in human occipitotemporal cortex. J Neurosci 35: 12977–12985. CrossRef Medline

Bréchet, L., Grivaz, P., Gauthier, B., & Blanke, O. (2018). Common recruitment of angular gyrus in episodic autobiographical memory and bodily self-consciousness. Frontiers in behavioral neuroscience, 270.

Brady N, Campbell M, Flaherty M (2004) My left brain and me: a dissociation in the perception of self and others. Neuropsychologia 42:1156–1161.

Burling, J. M., Kadambi, A., Safari, T., & Lu, H. (2019). The impact of autistic traits on self-recognition of body movements. Frontiers in psychology, 9, 2687.

Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E., & Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. Current biology, 16(19), 1905-1910.

Candidi, M., Stienen, B. M., Aglioti, S. M., & de Gelder, B. (2011). Event-related repetitive transcranial magnetic stimulation of posterior superior temporal sulcus improves the detection of threatening postural changes in human bodies. Journal of Neuroscience, 31(48), 17547-17554.

Chang, D. H., Troje, N. F., Ikegaya, Y., Fujita, I., & Ban, H. (2021). Spatiotemporal dynamics of responses to biological motion in the human brain. Cortex, 136, 124-139.s

Cisler, J. M., Bush, K., & Steele, J. S. (2014). A comparison of statistical methods for detecting context-modulated functional connectivity in fMRI. *Neuroimage*, *84*, 1042-1052.

Conson, M., Aromino, A. R., & Trojano, L. (2010). Whose hand is this? Handedness and visual perspective modulate self/other discrimination. Experimental Brain Research, 206(4), 449-453.

Coste, A., Bardy, B. G., Janaqi, S., Słowiński, P., Tsaneva-Atanasova, K., Goupil, J. L., & Marin, L. (2021). Decoding identity from motion: how motor similarities colour our perception of self and others. Psychological research, 85(2), 509-519.

David, N., Cohen, M. X., Newen, A., Bewernick, B. H., Shah, N. J., Fink, G. R., et al. (2007). The extrastriate cortex distinguishes between the consequences of one's own and others' behavior.Neuroimage, 36, 1004e1014.

Darda, K. M., & Ramsey, R. (2019). The inhibition of automatic imitation: A meta-analysis and synthesis of fMRI studies. NeuroImage, 197, 320-329.

De Bellis, F., Trojano, L., Errico, D., Grossi, D., & Conson, M. (2017). Whose hand is this? Differential responses of right and left extrastriate body areas to visual images of self and others' hands. Cognitive, Affective, & Behavioral Neuroscience, 17(4), 826-837.

de Gelder B (2006) Towards the neurobiology of emotional body language. Nat Rev Neurosci 7:242–249.

de Gelder B (2009) Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. Philos Trans R Soc Lond B Biol Sci
364:3475–3484.

Desmurget, M., Epstein, C. M., Turner, R. S., Prablanc, C., Alexander, G. E., & Grafton, S. T. (1999). Role of the posterior parietal cortex in updating reaching movements to a visual target. Nature neuroscience, 2(6), 563-567.

Devue, C., Collette, F., Balteau, E., Degueldre, C., Luxen, A., Maquet, P., & Brédart, S. (2007). Here I am: the cortical correlates of visual self-recognition. Brain research, 1143, 169-182.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. Experimental brain research, 91(1), 176-180.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. Science, 293(5539), 2470e2473. http://dx.doi.org/10.1126/science.1063414.

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

Downing, P. E., Chan, A. Y., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006). Domain specificity in visual cortex. Cerebral cortex, 16(10), 1453-1461.

Engelen, T., de Graaf, T. A., Sack, A. T., & de Gelder, B. (2015). A causal role for inferior parietal lobule in emotion body perception. cortex, 73(195), e202.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. Trends in cognitive sciences, 2(12), 493-501.

Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage, 15(4), 870-878.

Goldenberg, G., & Spatt, J. (2009). The neural basis of tool use. *Brain*, *132*(6), 1645-1655.

Graziano, M. S., Taylor, C. S., & Moore, T. (2002). Complex movements evoked by microstimulation of precentral cortex. Neuron, 34(5), 841-851.

Graziano, M. S. (2016). Ethological action maps: a paradigm shift for the motor cortex. Trends in cognitive sciences, 20(2), 121-132.

Grèzes J, Armony JL, Rowe J, Passingham RE. Activations related to “mirror” and “canonical” neurones in the human brain: an fMRI study, Neuroimage., 2003, vol. 18 (pg. 928-937)

Haaland, K. Y., Harrington, D. L., & Knight, R. T. (2000). Neural representations of skilled movement. *Brain*, *123*(11), 2306-2313.

Haaland, K. Y. (2006). Left hemisphere dominance for movement. *The clinical neuropsychologist*, *20*(4), 609-622.

Haxby, J. V., Connolly, A. C., & Guntupalli, J. S. (2014). Decoding neural representational spaces using multivariate pattern analysis. Annual review of neuroscience, 37, 435-456.

Hodzic, A., Kaas, A., Muckli, L., Stirn, A., & Singer, W. (2009). Distinct cortical networks for the detection and identification of human body. Neuroimage, 45(4), 1264-1271.

Hohmann, T., Troje, N. F., Olmos, A., & Munzert, J. (2011). The influence of motor expertise and motor experience on action and actor recognition. Journal of Cognitive Psychology, 23(4), 403–415.Return to ref 2011 in article

Iacoboni, M., Koski, L. M., Brass, M., Bekkering, H., Woods, R. P., Dubeau, M. C., ... & Rizzolatti, G. (2001). Reafferent copies of imitated actions in the right superior temporal cortex. Proceedings of the national academy of sciences, 98(24), 13995-13999.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. PLoS biology, 3(3), e79.

Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. Annual review of psychology, 60, 653-670.

Iacoboni, M. (2008). Mirroring people: The new science of how we connect with others. Farrar, Straus and Giroux.

Jackson, P. L., & Decety, J. (2004). Motor cognition: A new paradigm to study self–other interactions. Current opinion in Neurobiology, 14(2), 259-263.

Jeannerod, M. Visual and action cues contribute to the self–other distinction. Nat Neurosci 7, 422–423 (2004). https://doi.org/10.1038/nn0504-422

Jeannerod, M., & Pacherie, E. (2004). Agency, simulation and self-identification. Mind & language, 19(2), 113-146.

Jenkinson, M and Smith, S. A global optimisation method for robust affine registration of brain images. Medical Image Analysis, 5(2):143-156, 2001.

Jenkinson, M., Bannister, P., Brady, J., and Smith, S. Improved optimisation for the robust and accurate linear registration and motion correction of brain images. NeuroImage, 17(2):825-841, 2002.

Jokisch, D., Daum, I., & Troje, N. F. (2006). Self recognition versus recognition of others by biological motion: Viewpoint-dependent effects. Perception, 35(7), 911-920.

Kadambi, A., Xie, Q., & Lu, H. (2022). Motoric and intrinsic influences modulate self-recognition of actions. Under Review.

Kaplan, J. T., Aziz-Zadeh, L., Uddin, L. Q., & Iacoboni, M. (2008). The self across the senses: an fMRI study of self-face and self-voice recognition. *Social cognitive and affective neuroscience*, *3*(3), 218-223.

Keenan, J. P., Wheeler, M., Platek, S. M., Lardi, G., & Lassonde, M. (2003). Self-face processing in a callosotomy patient. European Journal of Neuroscience, 18(8), 2391-2395.

Koski, L., Iacoboni, M., Dubeau, M. C., Woods, R. P., & Mazziotta, J. C. (2003). Modulation of cortical activity during different imitative behaviors. Journal of neurophysiology.

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

Leung, H. C., & Cai, W. (2007). Common and differential ventrolateral prefrontal activity during inhibition of hand and eye movements. Journal of Neuroscience, 27(37), 9893-9900.

Leshinskaya, A., & Caramazza, A. (2014). Nonmotor aspects of action concepts. Journal of Cognitive Neuroscience, 26(12), 2863-2879.

Lieberman, M. D., Straccia, M. A., Meyer, M. L., Du, M., & Tan, K. M. (2019). Social, self,(situational), and affective processes in medial prefrontal cortex (MPFC): Causal, multivariate, and reverse inference evidence. Neuroscience & Biobehavioral Reviews, 99, 311-328.

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

Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., Wheelwright, S. J., Sadek, S. A., Suckling, J., ... & Baron-Cohen, S. (2010). Shared neural circuits for mentalizing about the self and others. Journal of cognitive neuroscience, 22(7), 1623-1635.

Loula, F., Prasad, S., Harber, K., & Shiffrar, M. (2005). Recognizing people from their movement. Journal of Experimental Psychology: Human Perception and Performance, 31(1), 210.

Macuga, K. L., & Frey, S. H. (2011). Selective responses in right inferior frontal and supramarginal gyri differentiate between observed movements of oneself vs. another. Neuropsychologia, 49(5), 1202-1207.

McLaren, D. G., Ries, M. L., Xu, G., & Johnson, S. C. (2012). A generalized form of context-dependent psychophysiological interactions (gPPI): a comparison to standard approaches. Neuroimage, 61(4), 1277-1286.

Molnar-Szakacs, I., & Uddin, L. Q. (2013). Self-processing and the default mode network: interactions with the mirror neuron system. Frontiers in human neuroscience, 7, 571.

Morin, A., & Michaud, J. (2007). Self-awareness and the left inferior frontal gyrus: inner speech use during self-related processing. Brain research bulletin, 74(6), 387-396.

Mumford, J., Turner, B., Ashby, G., & Poldrack, R. (2012). Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. Neuroimage, 59(3), 2636–2643. doi: 10.1016/j.neuroimage.2011.08.076

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. Current biology, 20(8), 750-756.

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. Neuroimage, 25(3), 653-660.

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

Orgs G, Dovern A, Hagura N, Haggard P, Fink GR, Weiss PH (2016) Constructing visual perception of body movement with the motor cortex. Cereb Cortex 26:440–449

Peelen, M. V., & Downing, P. E. (2005). Selectivity for the human body in the fusiform gyrus. Journal of neurophysiology, 93(1), 603-608.

Pereira, F., & Botvinick, M. (2011). Information mapping with pattern classifiers: a comparative study. Neuroimage, 56(2), 476-496.

Price, C. J., & Friston, K. J. (1997). Cognitive conjunction: a new approach to brain activation experiments. Neuroimage, 5(4), 261-270.

Platek, S. M., Wathne, K., Tierney, N. G., & Thomson, J. W. (2008). Neural correlates of self-face recognition: an effect-location meta-analysis. Brain research, 1232, 173-184.

Platek, S. M., Loughead, J. W., Gur, R. C., Busch, S., Ruparel, K., Phend, N., ... & Langleben, D. D. (2006). Neural substrates for functionally discriminating self-face from personally familiar faces. Human brain mapping, 27(2), 91-98.

Qin, P., Wang, M., & Northoff, G. (2020). Linking bodily, environmental and mental states in the self—A three-level model based on a meta-analysis. Neuroscience & biobehavioral reviews, 115, 77-95.

Raine, A. (1991). The spq: A scale for the assessment of schizotypal personality based on DSM-III-r criteria. *Schizophrenia Bulletin*. https://doi.org/10.1093/schbul/17.4.555

Roberts, R., Callow, N., Hardy, L., Markland, D., & Bringer, J. (2008). Movement imagery ability: development and assessment of a revised version of the vividness of movement imagery questionnaire. Journal of Sport and Exercise Psychology, 30(2), 200-221.

Rissman, J., Gazzaley, A., & D'Esposito, M. (2004). Measuring functional connectivity during distinct stages of a cognitive task. Neuroimage, 23(2), 752-763.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. Annu. Rev. Neurosci., 27, 169-192.

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

Roberts, R., Callow, N., Hardy, L., Markland, D., & Bringer, J. (2008). Movement imagery ability: Development and assessment of a revised version of the vividness of movement imagery questionnaire. *Journal of Sport and Exercise Psychology*. https://doi.org/10.1123/jsep.30.2.200

Sasaki, A. T., Okamoto, Y., Kochiyama, T., Kitada, R., & Sadato, N. (2018). Distinct sensitivities of the lateral prefrontal cortex and extrastriate body area to contingency between executed and observed actions. Cortex, 108, 234–251.

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

Soch, J., Deserno, L., Assmann, A., Barman, A., Walter, H., Richardson-Klavehn, A., & Schott, B. H. (2017). Inhibition of information flow to the default mode network during self-reference versus reference to others. Cerebral Cortex, 27(8), 3930-3942.

Sokolov AA, Zeidman P, Erb M, Ryvlin P, Friston KJ, Pavlova, MA. 2018. Structural and effective brain connectivity underlying biological motion detection. Proc Natl Acad Sci U S A. 115:E12034–E12042.

Sperry, R. W., Zaidel, E., & Zaidel, D. (1979). Self recognition and social awareness in the deconnected minor hemisphere. *Neuropsychologia*, *17*(2), 153-166.

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

Sugiura, M., Watanabe, J., Maeda, Y., Matsue, Y., Fukuda, H., & Kawashima, R. (2005). Cortical mechanisms of visual self-recognition. Neuroimage, 24(1), 143-149.

Tholen, M. G., Schurz, M., & Perner, J. (2019). The role of the IPL in person identification. Neuropsychologia, 129, 164-170.

Turk DJ, Heatherton TF, Kelley WM, Funnell MG, Gazzaniga MS, Macrae CN (2002) Mike or me? Self-recognition in a split brain patient. Nat Neurosci 5:841–842

Uddin, L. Q., Kaplan, J. T., Molnar-Szakacs, I., Zaidel, E., & Iacoboni, M. (2005). Self-face recognition activates a frontoparietal "mirror" network in the right hemisphere: an event-related fMRI study. *Neuroimage*, *25*(3), 926-935.

Uddin, L. Q., Rayman, J., & Zaidel, E. (2005). Split-brain reveals separate but equal self-recognition in the two cerebral hemispheres. *Consciousness and cognition*, *14*(3), 633-640.

Uddin, L. Q., Iacoboni, M., Lange, C., & Keenan, J. P. (2007). The self and social cognition: the role of cortical midline structures and mirror neurons. *Trends in cognitive sciences*, *11*(4), 153-157.

Uhlmann, L., Pazen, M., van Kemenade, B. M., Steinsträter, O., Harris, L. R., Kircher, T., & Straube, B. (2020). Seeing your own or someone else's hand moving in accordance with your action: The neural interaction of agency and hand identity. Human brain mapping, 41(9), 2474-2489.

Urgen, B. A., Pehlivan, S., & Saygin, A. P. (2019). Distinct representations in occipito-temporal, parietal, and premotor cortex during action perception revealed by fMRI and computational modeling. Neuropsychologia, 127, 35-47.

Urgesi, C., Candidi, M., Ionta, S., & Aglioti, S. M. (2007). Representation of body identity and body actions in extrastriate body area and ventral premotor cortex. *Nature neuroscience*, *10*(1), 30-31.

van Boxtel, J. J., & Lu, H. (2013). A biological motion toolbox for reading, displaying, and manipulating motion capture data in research settings. Journal of vision, 13(12), 7-7.

Van Den Bos, E., and Jeannerod, M. (2002). Sense of body and sense of action both contribute to self-recognition. *Cognition* 85, 177–187. doi: 10.1016/S0010-0277(02)00100-2

van Kemenade, B. M., Arikan, B. E., Kircher, T., & Straube, B. (2017). The angular gyrus is a supramodal comparator area in action–outcome monitoring. Brain Structure and Function, 222(8), 3691-3703.

Van Kemenade, B. M., Arikan, E. A., Podranski, K., Steinsträter, O., Kircher, T., & Straube, B. (2019). Distinct roles for the cerebellum, angular gyrus and middle temporal gyrus in action-feedback monitoring. Cerebral Cortex, 29, 1520–1531. https://doi.org/10.1093/cercor/ bhy048

van Veluw, S. J., & Chance, S. A. (2014). Differentiating between self and others: an ALE meta-analysis of fMRI studies of self-recognition and theory of mind. *Brain imaging and behavior*, *8*(1), 24-38.

Vocks, S., Busch, M., Grönemeyer, D., Schulte, D., Herpertz, S., & Suchan, B. (2010). Differential neuronal responses to the self and others in the extrastriate body area and the fusiform body area. Cognitive, Affective, & Behavioral Neuroscience, 10(3), 422-429.

Weiner, K. S., & Grill-Spector, K. (2011). Not one extrastriate body area: using anatomical landmarks, hMT+, and visual field maps to parcellate limb-selective activations in human lateral occipitotemporal cortex. Neuroimage, 56(4), 2183-2199.

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *Neuroimage*, *92*, 381-397.

Worsley, K. J. (2001). Statistical analysis of activation images. Functional MRI: An introduction to methods, 14(1), 251-70.

Wozniak, M. M., Schmidt, T. T., Wu, Y., Blankenburg, F., & Hohwy, J. (2021). Differences in working memory coding of biological motion attributed to oneself and others

Wurm, M. F., & Caramazza, A. (2021). Two 'what'pathways for action and object recognition. *Trends in cognitive sciences*.

Wurm, M. F., & Lingnau, A. (2015). Decoding Actions at Different Levels of Abstraction. Journal of Neuroscience, 35(20), 7727–7735. https://doi.org/10.1523/JNEUROSCI.0188-15.2015

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

Yovel, G., & O'Toole, A. J. (2016). Recognizing people in motion. Trends in cognitive sciences, 20(5), 383-395.

CHAPTER 4

Visual perception of social incongruencies: How interactions go awry

Abstract

Dyadic interactions can sometimes elicit a disconcerting response from viewers, generating a sense of “awkwardness”. Despite the ubiquity of awkward social interactions in daily life, it remains unknown what visual cues signal the oddity of human interactions and yield the subjective impression of awkwardness. In the present experiments, we focused on a range of greeting behaviors (handshake, fist-bump, high-five) to examine both the inherent objectivity and impact of contextual and kinematic information in the social evaluation of awkwardness. In Experiment 1, we analyzed participants’ verbal descriptions about greeting behaviors presented in raw videos. Participants showed consensus in judging awkwardness from greeting behaviors, and consistently used social and motor related words in their descriptions for awkward interactions. Experiment 2 employed advanced computer vision techniques to present the same greeting behaviors in three different display types. All display types preserved kinematic information but varied contextual information: (1) *Patch displays* presented blurred scenes comprised of patches (2) *Body displays* presented human body figures on a black background, (3) *Skeleton displays* presented skeletal figures of moving bodies. Participants rated the degree of awkwardness of greeting behaviors. Across display types, participants consistently discriminated awkward and natural greetings, indicating that the kinematics of body movements primarily drives awkwardness judgments. Multidimensional Scaling analysis based on the similarity of awkwardness ratings revealed two primary cues: social coordination and motor coordination. We conclude that the perception of awkwardness, while primarily inferred on the basis of kinematic information, is additionally impacted by the perceived social coordination underlying human greeting behaviors.

Introduction

In some social situations, dyadic interactions can elicit a disconcerting response from viewers—a sense of "awkwardness". For example, recent videos of U.S. President Donald Trump shaking hands with his Supreme Court nominee Neil Gorsuch (link) and the Prime Minister of Japan (link) generated a wave of discussions among both laypeople and experts about what exactly constitutes an "awkward" motoric social interaction (link).

Although difficult to pinpoint an exact definition, awkwardness is a subjective impression that can arise from many different cues, including a failure in executing motor behavior, misunderstood intentions, and conflicting personality traits. While perhaps amusing as a construct, when an interaction is perceived as awkward, either by participants or third-party observers, social goals and fluid communication are also likely to be impeded (Snyder, Tanke, & Berscheid, 1977), compromised more readily in clinical conditions associated with mentalizing difficulties, such as across the Autism-Spectrum (Heavy, Phillips, Baron-Cohen, and Rutter, 2000). Therefore, to sufficiently probe an underlying, underexamined, and heterogenous construct like "awkwardness," important questions emerge: are people idiosyncratic in their perceptions of awkwardness in dyadic interactions, or are people in general agreement? If people agree with each other in perceiving awkwardness, is it possible to pin down the visual characteristics contributing to the impression of awkwardness? What visual features signal the oddity of a dyadic interaction? The present paper explores these questions through examining judgments of awkwardness conveyed through human social greeting interactions.

Two research fields provide relevant knowledge about perceiving social attributes from visual input. Research on biological motion perception provides clear evidence that, when human actions are reduced to moving dots located at key joints (Johansson, 1976), human observers make reliable attributions of social properties such as deception (Runeson & Frykholm, 1983), intention (Hohmann, Troje, Olmos, & Munzert, 2011), affect (Pollick, Paterson, Bruderin, & Sanford, 2001), sex (Johnson and Tassinary, 2005), and personality traits (Borkenau et.al., 2004). Additionally, these studies highlight the importance of kinematic information in inferring the social properties of human actions.

A separate line of work in person perception has focused on a different but equally important question, the role of visual context in social evaluation. Humans live and operate within rich contextual environments. Human observers have been shown to use static images of faces to make reliable attributions of visually ambiguous social properties such as sexual orientation (Rule & Ambady, 2008; Rule et al., 2008), political identity (Rule & Ambady, 2010), and personality traits (see Todorov, Said, & Verosky, 2011 for a review). Related research has demonstrated that other aspects of visual context, such as race (Alter, Stern, Granot, & Balcetis, 2016), attire (Freeman, Penner, Saperstein, Scheutz, & Ambady, 2011) and scene background (Freeman et al., 2013) also influence the attribution of visually-based social properties. From these categorizations, visual context is multifaceted and wide-ranging. Hence, context could refer to something as low-level as the human voice which contextualizes a face (Freeman & Ambady, 2011) or as high-level as the physical scene in which a human is grounded.

Through integrating the two separate, but closely linked research fields, we aimed to provide a novel methodology to examine the role of visual context and human kinematic information in the perception of awkward greeting behaviors. Research in biological motion perception has shown that people do reliably use human kinematics to attribute social properties to human actions. However, daily encountered "awkward" social greetings, such as those conveyed through missed high-fives and handshakes, have not received sufficient empirical attention. Modern advances in deep learning algorithms make it possible to systematically segregate motor and contextual information, by segmenting body movements and extracting skeleton displays from raw action video recordings. To this end, we delineated varying aspects of visual context (described in Experiment 2), through presenting highly naturalistic and contextualized human interactions in different display types, to parametrically examine the contribution of contextual and kinematic information.

To assess whether people are in general agreement in their perceptions of awkwardness in dyadic interactions, we used both free response (Experiment 1) and rating (Experiment 2) paradigms. In Experiment 1, participants were presented with a variety of daily encountered greeting behaviors in raw videos and asked to identify whether each greeting behavior appeared awkward. If the video was categorized as awkward, participants were asked to describe why this categorization was made. The text analysis based on free responses allowed us to explore whether people consistently attend to certain social or motor cues in the raw videos to identify the presence of an awkward interaction.

Experiment 2 aimed to further examine the interpretability of the free response results by asking for subjective ratings on the experimentally manipulated stimuli. We employed advanced computer vision techniques to generate the stimuli of greeting actions presented in three different video display types. These displays consisted of dyadic interactions and varied the amount of visual context presented in the stimuli. To parameterize the amount of visual information, we broadly divided visual context into four main categories: body structure (i.e., joints and height), body morphology (i.e., width, shape, gender), actor identity (i.e., skin tone, coarse facial features, attire), and scene depiction (i.e., physical background- indoor versus outdoor scenes). We characterized kinematics as the information provided by body movements of the actors involved in each greeting action. Across the display types, we maintained the kinematic information of body movements, while removing particular categories of contextual information. For instance, in one type of display (discussed in more detail below), we removed the scene and actor identity information, but maintained body morphology and structure characteristics. In another display type, consisting of the sparsest visual information, we removed the scene, actor identity, and body morphology information, while just preserving the body structure. Through incorporating these categorizations of visual context, we examined the independent contribution of human kinematic information to awkwardness judgments as impacted by the gradual mitigation of contextual information in the different display types. To further elaborate, we describe the visual context provided by three different display types below:

The first type, *patch displays*, presented blurred scenes and featured the most amount of visual context of our three display types. Specifically, the patch displays preserved all four components of our broadly defined criteria of visual context: scene depiction, actor identity, body morphology, and body structure. They offered rich cues about the settings in which greeting actions occurred including the visual background, objects in the scene, and other actors not involved in greeting actions. They offered cues about actor identities including skin tone, coarse facial features (e.g., a separation of face area and hair), and attire. Additionally, they preserved body morphology information such as body shape and gender, as well as structural body information (joints and height).

The second type, *body displays*, presented colorful human body figures on a black background and featured less visual context than did patch displays. Specifically, both scene information (the physical setting) and actor identity information (skin tone, coarse facial features, attire) were removed, while preserving body morphology (cues about coarse body shape, such as width, height, and gender) and body structure (joints and height). Note that while the physical scene of the action was eliminated, sparse cues about other actors not involved in greeting behaviors (occasionally displaying body parts of background actors) remained.

The third type, *skeleton displays*, presented white skeleton figures resembling human body structure against a black background. The skeleton displays featured the least visual context of our three display types, preserving only the structural body information of the main actors in the display. Specifically, the skeleton displays included no cues about the physical setting in which greeting actions occurred, nor actor identity information such as skin tone and coarse facial expression, nor cues about body shape, such as body width and gender. Therefore, only body height and joint information was presented, depicted by stick-figures of the main actors involved in greeting behaviors.

Notably, all three display types held human kinematics constant as they were generated from the identical raw video recordings, but each display included different categories of visual context. This key experimental manipulation allowed us to compare awkwardness ratings of greeting behaviors across a range of different naturalistic actions, in order to examine the relationship between visual context and kinematic information in the social evaluation of greeting behaviors. To underscore, these display types were chosen due to their inherent similarity in the presentation of human body structure and kinematics, while systematically varying key components of visual context: body structure (consistent across displays), body morphology, actor identity, and background scene information.

Finally, Experiment 2 also attempted to clarify what key visual features might serve as cues to the perception of awkwardness. To this end, we conducted a multi-dimensional scaling analysis (MDS) based on participant awkwardness ratings to recover the two-dimensional psychological space in accordance to the difference of perceived awkwardness judgments amongst greeting behaviors. The interpretation of the two primary dimensions could provide convergent evidence to reveal the characteristics of free responses reported in Experiment 1 when people are explicitly asked to describe awkwardness.

Experiment 1

A free-response study was conducted to measure the perception of awkwardness in greeting behaviors from raw video recordings. Here, participants viewed videos of awkward and natural greeting behaviors and subsequently categorized the video as awkward or natural by providing written descriptions of the social interaction. The present experiment also explored features of the semantic descriptions judging awkward interactions, and whether the contribution of social and motor cues indeed signified the presence of an awkward interaction.

Methods

Participants

30 participants (m = 9, f = 21) were recruited from the University of California, Los Angeles (UCLA) Psychology Subject pool. Participants provided informed consent, as approved by the UCLA Institutional Review Board (#16-001879) and were given course credit for their participation.

Stimuli & Apparatus

Participants were tested in a dark, quiet room 76.2 cm from the display. Monitor width and height was 53.1° x 40.7°. Thirty-four videos from YouTube (see Appendix A for links) were selected to capture a variety of greeting behaviors ranging from awkward and natural social interactions. The videos varied in length (2 to 27 seconds, $M = 6.36$, $SD = 4.47$) and context, and were selected according to two important reasons. First, previous studies on interpersonal coordination of joint actions have largely categorized actions as either deliberate (such as high-fiving) or unnoticed (such as bumping into each other; e.g., Schmidt, Fitzpatrick, Caron, & Mergeche, 2011). By including videos that encompass both types of greeting actions, we were able to include a more variable set of interactions. Secondly, the selected videos depicted different degrees of awkwardness. The range of variability allowed for participants to appraise social situations that were more relevant and encountered on a daily basis, as awkward interactions could occur from different types of greeting behaviors in many situations (e.g., multiple individuals, varied contextual settings, greeting styles). Hence, the stimuli of awkward greetings are more heterogeneous than natural greeting behaviors. Given the limited number of psychological studies that use naturalistic videos to examine and operationalize awkwardness from daily interactions, the key perceptual signals signifying a heterogenous construct, such as awkwardness, remain unknown. Therefore, to cover a large range of awkward greeting behaviors, we included a greater number of videos featuring awkward greeting behaviors (24) than videos featuring natural greeting behaviors (10). Finally, videos including text describing an awkward interaction (e.g., video 22:"Trump's awkward interactions with world leaders") were cropped in order to not bias participant responses with the inclusion of text.

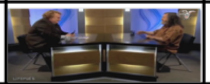
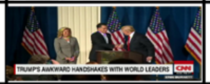
Procedure

Participants were presented with a randomized order of 34 raw videos selected from YouTube. For each of the 34 videos (and consequently after each trial), participants were asked to categorize the greeting behavior in the video as "awkward" or "not awkward". Note that "awkward" and "not awkward" were not defined to the participant, thus allowing the participant to use their own criteria. If the participants labeled the video as awkward, they were instructed to write a brief verbal description explaining why the greeting behavior in the video appeared awkward. Participants were not given a time limit to write their descriptions. No sound was provided to the participants during video presentation.

Results

In the first step, the proportion of participants identifying greeting behaviors as awkward was reported for each video (see Figure 25, column 5). The response proportion was used to classify each video as natural or awkward in the remaining paper. Each video was classified as natural if the mean proportion of participants categorizing it as awkward was less than .50, and a video was classified as awkward if this mean proportion was greater than .50. All 24 videos identified by the experimenters as awkward showed a mean proportion greater than .50. To reliably assess whether people were in general agreement in judging awkwardness from human interactions, a Spearman-Brown corrected split-half reliability coefficient indicated high internal consistency (r =.850) between participants for video awkwardness ratings.

Videos classified as Awkward				
Video ID	Frame from Video	Social Words	Motor Words	Proportion Awkward
1		26	32	1.00
2		18	11	0.97
3		29	19	0.97
4		20	22	0.93
5		38	27	0.93
6		17	29	0.93
7		25	19	0.93
8		25	24	0.93
9		41	9	0.90
10		17	21	0.90
11		26	16	0.90
12		25	23	0.90

Videos classified as Awkward				
Video ID	Frame from Video	Social Words	Motor Words	Proportion Awkward
13		11	21	0.87
14		16	12	0.87
15		28	11	0.87
16		56	9	0.87
17		19	16	0.83
18		27	20	0.83
19		19	9	0.80
20		18	16	0.77
21		17	13	0.77
22		8	25	0.77
23		21	13	0.77
24		18	5	0.73

Videos classified as Natural

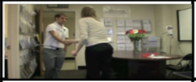
Video ID	Frame from Video	Social Words	Motor Words	Proportion Awkward
25		24	1	0.43
26		2	2	0.40
27		5	0	0.37
28		11	12	0.33
29		5	0	0.30
30		5	4	0.23
31		0	0	0.17
32		1	1	0.13
33		1	0	0.07
34		0	0	0.03

Figure 25. Examples of video classified as awkward (top left and right) and natural (bottom) videos including: key frame, corresponding frequency of social and motor word descriptions (generated from participants' written descriptions), and proportion of awkward responses (>.5 indicates that the greeting action in the video was perceived as being awkward, <.5 indicates that the perception was natural behavior.)

In the second step, we analyzed written descriptions of videos for the greeting behaviors that participants categorized as awkward. Written descriptions from all the participants were merged into one file in order to identify the high-frequency words including nouns, verbs, adverbs and adjectives. For reference, sample descriptions for the awkward videos included "*This was awkward because the man in the middle attempted to shake hands with two people at once, using his left hand for another person's right hand*" and "*The two men gestured but it was small enough* *that the other person did not catch on fast enough so they were almost playing footsie with their hands."* The description file was entered into Textalyser (http://textalyser.net), an online software that provides a ranking of the most frequently occurring words used in a body of text. After excluding words with fewer than three characters and numerals, Textalyser returned the 200 most frequent words in the entire set of participants' written descriptions. Of those 200 words, the first two authors selected a subset of words that consisted of verbs, adverbs, and adjectives and that excluded nouns and redundant words like "awkward" or "handshake" or words in phrases like "fist bump" or "high five". From that subset, the authors selected words that were either motor-related or social-related. Motor-related words expressed actions predominantly related to motor behavior (e.g. "pull", "grab", "reach") or their properties (e.g. "towards"). Social-related words had a wider range and included words related to mental attributes (e.g. "try", "want", or "confuse"). We also categorized social-related words as those whose use indicates social knowledge about appropriate greetings (e.g. "kiss", "long", "far"). For example, "far" is considered a social word because there appeared to be an ideal distance between two people based on inherent social knowledge (too "far"). See the full list in Appendix B, as shown in the word cloud display (http://worditout.com) in Figure 26.

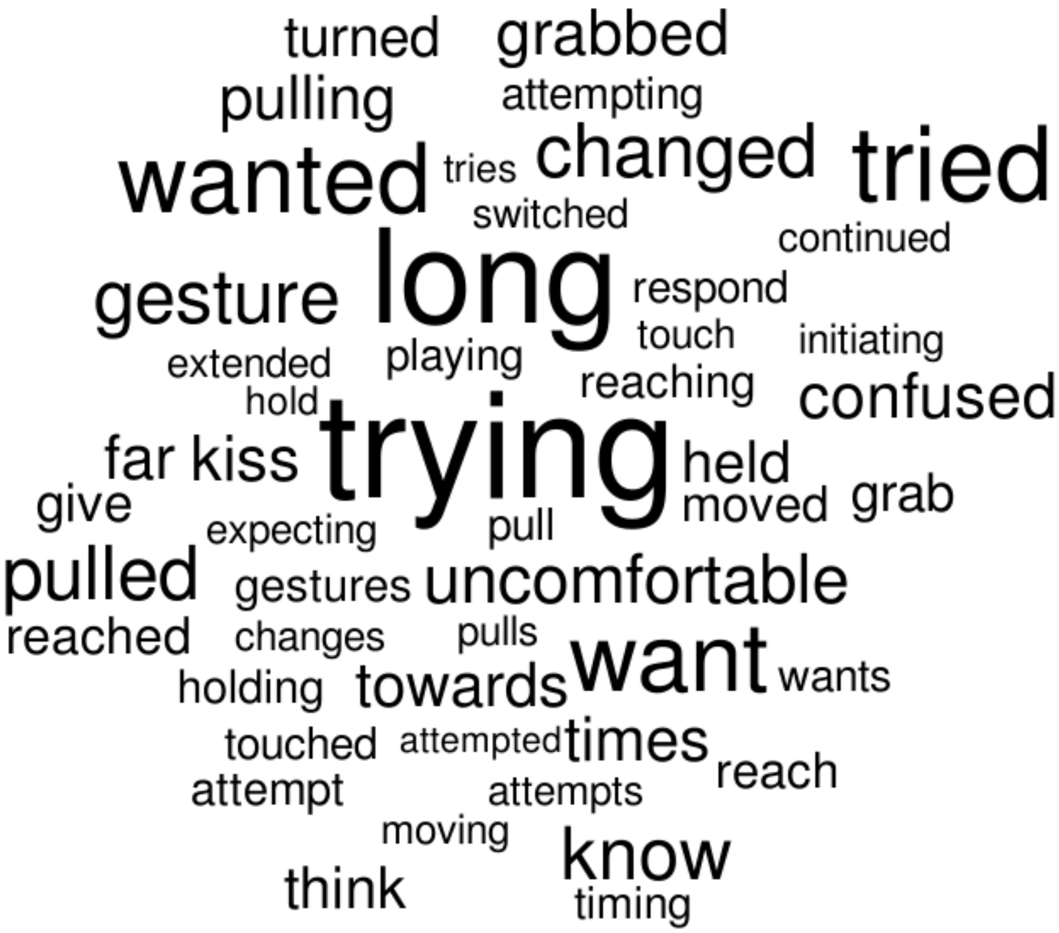


Figure 26. Word cloud displays both social-related and motor-related words used to freely describe awkward handshakes in Experiment 1. Font size corresponds to relative frequency.

The frequencies of the selected social-related and motor-related words in participants' written descriptions were then found for each of the 34 videos, individually (see Figure 25, columns 3 and 4). Note that when a social-related or motor-related word was a verb (e.g., try), all instances of its alternative forms (e.g. tried, tries, trying) were counted as instances of that word. Videos showed different ratios between the number of motor-related words and the number of social-related words in written descriptions. For example, for Video 1 in which Donald Trump aggressively attempts to pull a somewhat stiff and reluctant Neil Gorsuch closer and closer to him, participants' descriptions included the most motor-related words (33). For Video 16 in which a man seemingly reaches for a handshake from a woman and then attempts to kiss the woman's hand, participants employed the most number of social-related words (56) in written descriptions.

A Pearson correlation coefficient was computed to assess the relationship between the proportion of participants categorizing each video as awkward and the combined number of social-related and motor-related word descriptions for each video. We found a significant positive relationship, $r = .59$, $p = .002$, observed power = .965, suggesting that the perception of awkwardness is consistent with verbal descriptions of the videos. Additionally, while there was no significant relationship between the number of social words used in the descriptions and awkward proportion, there was a significant relationship between the number of motor words and awkward proportion ($r = .50$, $p = .012$, observed power = .863), suggesting the possible strong influence of motor cues on awkwardness judgments.

Collectively, the results of Experiment 1 revealed that people are qualitatively able to judge awkwardness from human interactions. Importantly, this ability is likely reliant on both social and motor cues, as revealed by the prevalence of social- and motor-related words in the semantic description of awkward greetings.

Experiment 2

Experiment 1 served as an exploratory study, demonstrating that participants are not idiosyncratic in their perceptions of awkwardness when viewing raw videos of greeting behaviors. Experiment 2 was designed to systematically measure awkwardness judgments by varying visual information provided in the display and to assess the contributions from the kinematics of body movements and the visual context in perceived awkwardness of greeting behaviors.

Method

Participants

Sixty-six participants (f = 49, m = 17, mean age = 20.62) were recruited from the University of California, Los Angeles (UCLA) Psychology Subject pool. All but two participants had lived in the United States for at least five years and were fluent English speakers. The sample size was determined in accordance with a previous study on individual differences in biological motion perception (van Boxtel, Peng, Su, & Lu, 2016). Participants provided informed consent, as approved by the UCLA Institutional Review Board (#16-001879) and were given course credit for their participation.

Stimuli and Apparatus

The identical 34 videos from Experiment 1 were also used to generate the stimuli in Experiment 2. The visual stimuli were presented on the center of the screen at the size of 250 x 250 pixels using MATLAB (R2017a; Mathworks, Natick, MA). The greeting behavior stimuli were manipulated by one of the three display types: patch, body, or skeleton display (see Figure 27 for sample image frames).

Patch display. The superpixel algorithm segmented raw images into patches by grouping pixels with similar brightness, color, and texture in the local regions of images. The output of this algorithm displayed patches of the locally consistent pixels with similar colors or gray levels (Ren & Malik, 2003). As shown in the second column of Figure 27, MATLAB's "superpixel" function was used to process the 34 raw videos to generate the patch display as a blurred visual scene. Since the patch display included blurred backgrounds and other people/objects in the display, the stimuli in the patch display provided the most contextual information in the visual scene.

Body display. We used a deep learning model to segment human bodies in the video. The model, RefineNet, exploits visual features at multiple levels of abstraction for high resolution semantic segmentation. The model classified the pixels of each video frame as belonging to a human form or not and further classified pixels corresponding to humans into the following body regions: head, torso, upper arm, lower arm, upper leg, and lower leg (Lin, Milan, Shen, & Reid, 2017). RefineNet processed the 34 raw videos to present colored human figures against a black background, as shown in the third column of Figure 25.

Skeleton display. We used a multi-person pose estimation algorithm (Cao, Simon, Wei, & Sheikh, 2016) to estimate the location of key body joints in videos. This deep learning model detects body parts and is robust against occlusion and viewpoints. Based on the inferred joint locations, skeleton figures were generated using BioMotion toolbox (van Boxtel & Lu, 2013) to extract the kinematic movement from the raw 34 videos, as shown in the fourth column of Figure 25. The white skeleton was displayed against a black background. At the time of conducting the experiment (July 2017), the pose estimation algorithm did not include options to infer the hands of the actors, and only provided estimation of joint coordinates up until the wrist for arms. Therefore, the skeleton actors did not directly touch in the stimuli. Participants were informed of this display feature in order to minimize the surprisal that the two hand-shaking actors did not touch the other person's skeleton. Additionally, the model occasionally failed to extract lower body parts, largely due to the similar color of pants as the background, or the missing parts (such as lower legs) occluded by objects (such as a table) in the YouTube videos. To correct for this, and maintain consistency across videos, a gray rectangular occluder was displayed at the bottom of the screen, which covered missing body parts.

Procedure

Participants were randomly assigned to view one of three display types (i.e., patch, body, skeleton displays). Participants first viewed a sample video (which was manipulated by the assigned display type and was not included in the experimental test trials), in order to gain familiarity with the display type. Following exposure to the sample video, participants began the experiment. In each trial, participants were asked to rate the degree of awkwardness of the greeting behavior in the video stimulus on a 6-point scale from 1 (surely natural), 2 (probably natural), 3 (guess natural), 4 (guess awkward) and 5 (probably awkward) to 6 (surely awkward). The subjects were not imposed with a time limit during the rating period and no sound was provided to the participants from the videos. The experiment consisted of 34 trials with randomized order and lasted around 30 minutes.

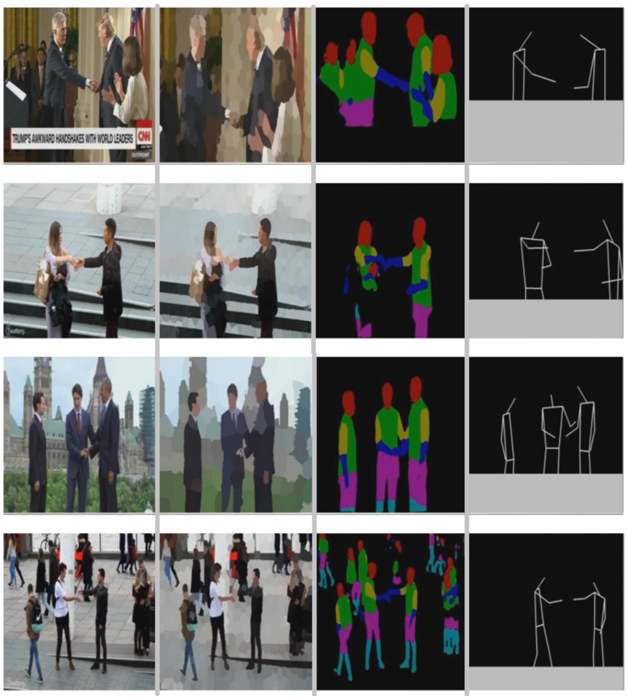


Figure 27. Example of stimuli in Experiment 2. Column 1, raw YouTube video frames; Column 2, patch display processed by superpixel algorithm; Column 3, body display processed by RefineNet model; Column 4, skeleton display converted from raw video.

Results

Mean human ratings for the three display types in Experiment 2 were significantly correlated with the proportion of awkward responses for the raw video recordings in Experiment 1 ($r = .92$ for patch, observed power = .999, $r = .83$ for body, observed power = .999, and $r = .84$, observed power = .999 for skeleton, *ps* <.001), suggesting that participants were in general agreement about categorizing awkward or natural interactions across all displays.

Next, we conducted a mixed ANOVA with one within-subjects factor, activity normality (awkward vs. natural greetings) and one between-subjects factor, display type (patch vs body vs skeleton). The activity normality of videos was determined by the proportion of responses in Experiment 1 that classified the videos as awkward (proportion >.5) or natural (proportion < .5). The average ratings for awkward videos and natural videos were used as dependent variables in the ANOVA analysis. As shown in Figure 28, a main effect of activity normality was revealed to show higher ratings for awkward videos than for natural videos, $F(1,63) = 539.100$, $p < .001$, $\eta_p^2 = .895$. We also found a significant two-way interaction effect between activity normality and display types ($F(2,63) = 11.096$, $p < .001$, $\eta_p^2 = .260$). Specifically, the impact of display type on ratings was not found for natural videos (patch vs body; patch vs skeleton, *p*s > .05). However, for awkward videos, participants yielded greater awkwardness ratings for the patch display than for the other two displays (patch vs. body, $p = .003$; patch vs. skeleton, $p < .001$, with Bonferroni correction). Since both the skeleton and body displays remove a majority of the contextual information (such as background, colors, race), the lower awkwardness ratings in these two displays reveal the strong influence of contextual information in judgments when observing awkward human interactions. This impact of display type on ratings for awkward videos is also consistent with the observations that awkwardness perceived in some videos differs depending on the display type. For example, the famous video of President Donald Trump shaking hands with his Supreme Court nominee (Video 1) was ranked highly awkward in the body display (rank #5; ranging from 1, most awkward, to 34, least awkward). But in the skeleton and patch display, people gave lower awkwardness ratings (#12 for skeleton, #8 for patch). Another example is Video 7. This video involves a scene where a person is intentionally avoiding a second person's high-five, fist-bump, and hug. Here, the video is consistently rated as awkward in the displays with increased contextual information (as rank #4 in the patch display and rank #5 in the body display). However, the skeleton display of the video was no longer judged high on awkwardness (rank #13). This discrepancy may result from the lack of contextual information (such as male/female, scenery) in the skeleton display, and the missing subtle kinematic information (such as missing hand movements). Participants' written descriptions for Video 7 match these observations, with 65% more social words than motor words.

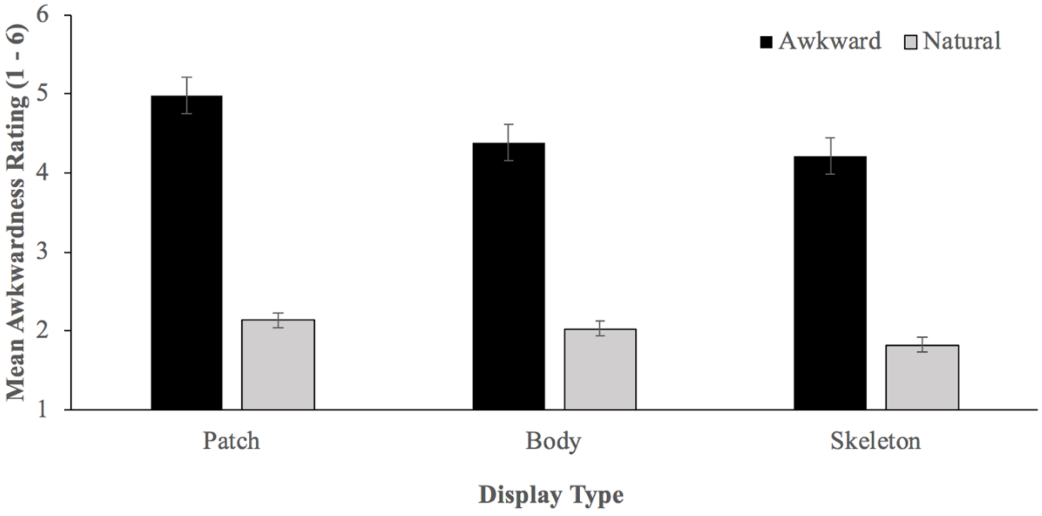


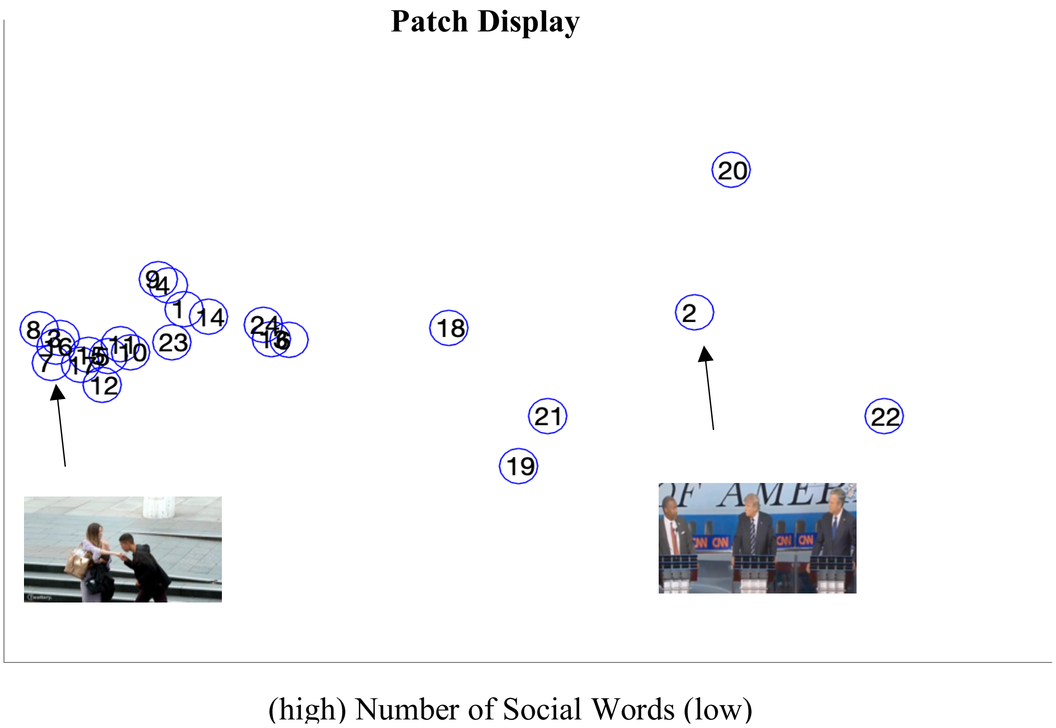
Figure 28. Mean awkwardness ratings for awkward and natural videos as a function of three types of displays. The error bars indicate standard error of the means.

Since our experiment included human interaction videos with different durations, we conducted a multiple regression analysis to assess whether video duration significantly accounted for variability in participants' awkwardness ratings. We measured participants' awkwardness ratings for the patch, body, and skeleton displays through utilizing two predictors: the proportion of awkwardness for each video (determined in Experiment 1) and video duration (measured in seconds). For all three display types, the proportion of participants classifying each video as awkward statistically significantly predicted awkwardness ratings. For the patch display, a significant regression equation was found $F(2, 31) = 67.929$, $p < .001$, with an $R^2 = .814$. However, there was no relationship between video duration and mean awkwardness ratings for the patch display ($\beta = .094$, n.s.), while the proportion of participants classifying each video as awkward served as a significant predictor ($\beta =.861$, p < .001). When mean awkwardness ratings for each video in the body display was predicted, a significant regression equation was found $F(2, 31) = 27.514$, $p < .001$. The overall model fit was $R^2 = .739$. Similarly, the proportion of participants classifying the video as awkward was a significant predictor ($\beta =.796$, p < .001), while there was no linear relationship between video duration and mean awkwardness ratings for the body display ($\beta = .136$, n.s.). For the skeleton display, a significant regression equation was also found $F(2, 31) = 37.420$, $p < .001$, with $R^2 = .707$. Additionally, the proportion of participants classifying the video as awkward served as a significant predictor ($\beta =.773$, p < .001), while there was no significant relationship between video duration and mean awkwardness ratings for the body display ($\beta =.146$, p < .001), consistent with model results from the patch and body display. Therefore, the varied video duration did not appear to influence participant's awkwardness judgments. Collectively, these findings indicate that participants reliably categorized videos as awkward and natural regardless of display type, highlighting the importance of kinematics to awkwardness judgments. Additionally, the proportion of participants categorizing each video as awkward served as an important predictor of awkwardness ratings for each of the display types, accounting for most of the variability in awkwardness ratings for each of the video display types. Finally, the degree to which the proportion of awkwardness correlated with ratings for each display type still varied according to the amount of contextual information present in each display type, with the most variability in awkwardness judgments in the patch display accounted for by the participants' proportion of awkwardness.

To better understand a psychological space underlying the awkwardness judgements, we conducted a Multidimensional Scaling (MDS) analysis (Kruskal & Wish, 1978) to explore what psychological dimensions play key roles in determining participants' ratings for awkwardness. We first included all 34 videos in the MDS analysis, however the MDS results appeared to cluster all the awkward videos in similar locations to separate from the natural videos, unhelpful for visualizing the basic features sensitive to the different degrees of awkward behavior. Hence, in the final MDS analysis, we only included ratings for the 24 awkward videos identified in Experiment 1. We computed the Euclidean distance between any pairs of ratings for awkward videos to generate the distance matrix for the 24 videos for each display type. Smaller distances reflected that the pair of actions were judged with similar awkwardness ratings across subjects. The 24 X 24 distance matrix was the input for the nonmetric MDS to project the pairwise distances of awkwardness ratings to a two–dimensional space.

The resultant space of MDS analysis was a two-dimensional psychological space with r^2 = .94, .92, .83; and stress = .07, .10, .17 for the displays of patch, body, and skeleton, respectively. Across all display types, we found a significant correlation between the horizontal coordinates of videos and the number of social words present from descriptions of the videos in Experiment 1 (for patch display, $r = -0.488$, $p = .015$, observed power = .686; body display, $r = -.562$, $p = .004$, observed power = .514; skeleton display, $r = -.611$, $p = .002$, observed power = .903). These correlation results were consistent with individual observation of clusters in Figure 29, which revealed that actions exhibiting a higher degree of social incoordination (with more social word descriptors) were consistently located at the left end of the resultant psychological space in the MDS result plot. For example, Video 16 was consistently located on the left side of the MDS result plot, wherein a man attempted to kiss (in lieu of handshaking) a young woman who avoided the body contact. This action involves a high degree of social incoordination due to the lack of engagement of the female actor, and a general violation of social heuristics (kissing instead of handshaking). On the other hand, awkward videos with a lower degree of social incoordination (determined by their lower degree of social descriptors from Experiment 1) were clustered on the right side. For example, Video 2, showing President Trump catching and vigorously shaking Ben Carson's hand at a presidential debate, was rated as highly awkward for the raw video. After removing the identity information, the handshaking videos in all the three displays were located on the right side of the MDS space, due to the high degree of social coordination in their interaction. Hence, the horizontal dimension in the psychological space of awkwardness judgment reveals the perceived degree of social incoordination, an overall impression of how well the two actors coordinate their social interaction in the greeting behavior.

The first dimension (horizontal) accounted for most of the variability in the awkwardness judgments across the videos**:** 86%, 79%, and 50% of the variance for the displays of patch, body, and skeleton, respectively. In contrast, the vertical dimension accounted for less variability, as 44%, 21%, and 16% of the variance for the three displays. The interpretation of the vertical dimension is not as clear as the horizontal dimension. Since the number of social words for each video from experiment 1 correlated with the horizontal axis, we explored whether the vertical axis correlated with the number of motor words for each video from experiment 1, but found no relationship across all display types. However, we noticed a possible relation with the touching duration for the skeleton display. For each video, we estimated whether "touching", defined by the distance between the two wrist points, was less than the average lower arm length. We found that the vertical locations of videos in the psychological space showed a marginal correlation with touching duration in the greeting behaviors in the skeleton display, $r = .390$, $p = .060$, observed power = .502. We conjecture that, when only body kinematics are available in the visual input, physical contact may serve as an important cue for judging the actions, likely related to internal knowledge regarding the appropriate duration of touching in the present greeting behaviors. However, when more contextual information is available in the interaction, other contextual cues, aside from touching duration, may more strongly contribute to awkwardness judgments, as we did not find a significant relationship between touching duration and the vertical axis for the more contextually rich patch and body displays.



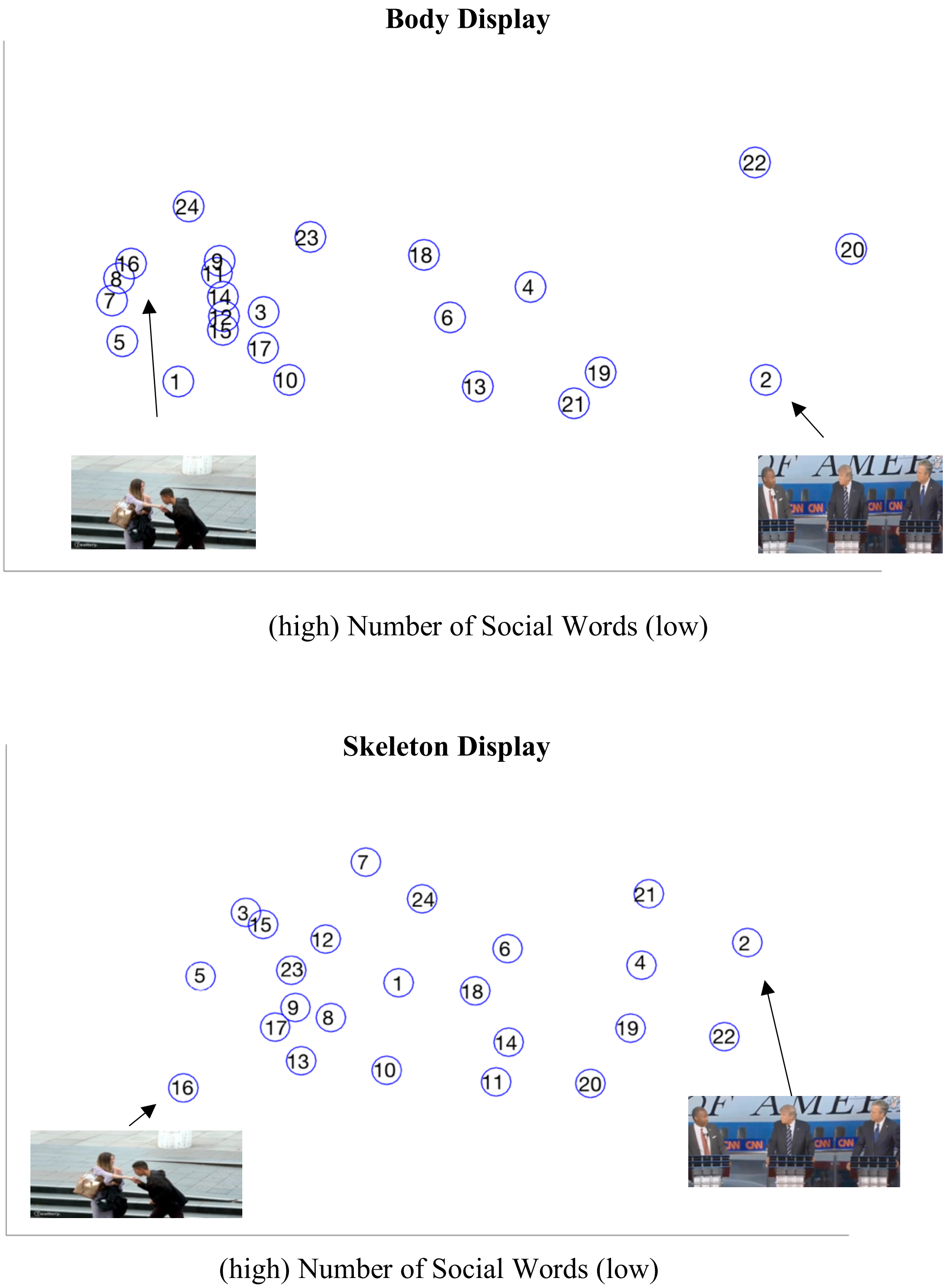


Figure 29. Results of the psychological space from MDS analysis for Patch Display (top), Body Display (middle), and Skeleton Display (bottom). Video 16 (boy kisses girl's hand) and 2 (Donald Trump's aggressive handshake-grab) are representative of videos consistently appearing in the similar horizontal location across display types, with Video 16 having more social descriptors and Video 2 having less.

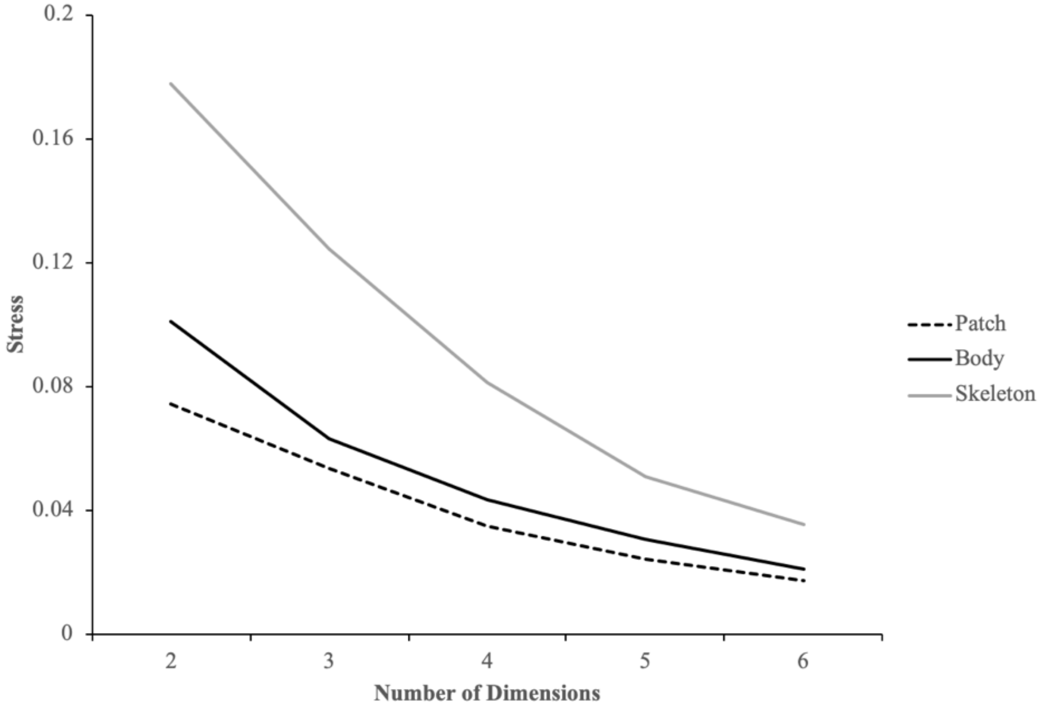


Figure 30. Stress plot of MDS representation for each of the three display types. A two-dimensional space was selected as the best choice since there was only a marginal improvement in stress with increased dimensionality.

Discussion

Humans encounter and experience awkward social interactions on a daily basis, yet previous research has not explored this complex social construct, nor investigated the contribution of kinematics to these judgments. Largely absent, is an interdisciplinary “person construal” approach (Freeman and Amabady, 2011), relating lower-level perceptual mechanisms (e.g., kinematics) with attributions of higher-level social judgments of awkwardness. Therefore, in two experiments, we examined whether awkwardness is inherently subjective, or whether there exists a more objective, streamlined criterion that humans use to reliably categorize awkwardness. Experiment 1 served as a preliminary study examining whether awkwardness can be reliably judged from greeting behaviors, and how this ability related to the presence of social and motor cues. Utilizing the written descriptors from Experiment 1, we rank-ordered the videos based on the proportion of participants classifying each interaction as awkward. In Experiment 2, we manipulated the amount of visual information while holding body kinematics constant. Together, the present experiments revealed that participants were systematically able to judge awkward behaviors across all three display types, underscoring the potential importance of human kinematics to social interpretations and its key role in signaling awkward behavior. This ability also appeared to account for the presence of contextual information (body morphology, actor identity, and scene depiction), as revealed by significantly greater awkwardness ratings for the patch display (with the highest degree of contextual information) than the body or skeleton displays.

To compare these results with individual cases, we examined particular videos judged with higher awkwardness ratings. We found that the video of President Donald Trump shaking hands with Neil Gorsuch (Video 1) was rated consistently higher (i.e., more awkward) than many of the other videos, even when visual cues to identity were removed, or reduced (as in the skeleton and body displays), suggesting that the perception of President Trump's awkward handshaking may primarily be attributed to his motoric "awkward" behavior (barring external influences of contextual differences). While visual identity generally appears to play a predominant role in social judgments (e.g., knowledge of President Trump) as well as in action recognition (e.g., Ferstl, Bulthoff, & de la Rosa, 2016), even observing the kinematics of human movements is a viable tool to make high-level social judgments of interactions. In fact, participants' written descriptions on the raw video corroborated this finding, by including double the amount of motor descriptions than social descriptions in their awkwardness descriptions.

Previous research has found that certain human motor cues play an important role in the detection of threatening actions (van Boxtel & Lu, 2011, 2012), perception of social interactions (Thurman & Lu, 2014), emotion perception from actions (Roether, Omlor, Christensen & Giese, 2009) and action discrimination (van Boxtel & Lu, 2015). What specific cues impact the perception of awkwardness in social interactions? To probe the underlying psychological dimensions of awkwardness judgments in greeting behaviors, Multidimensional Scaling (MDS) analysis revealed two important candidates: social coordination and touching duration (overall length of handshake/greeting). Social coordination accounted for most of the variability in judging awkwardness. Here, we define social coordination as the degree to which individual actions are in accordance not only with greeting behavior norms (e.g., shaking hands instead of kissing), but also with the physical setting in which the interaction takes place. Individual cases, such as President Trump catching and shaking Ben Carson's hand, demonstrate how (although ranked awkward) the interaction does not violate social coordination within the presidential debate setting (clustered on the right, or congruent, side). However, in cases of the first-time meeting with another individual in a public setting, a violation of social appropriateness in American culture is likely to occur when a stranger attempts to grab and kiss another individual's hand (as seen in Video 16 clustered on the opposite side in psychological space).

Through visual inspection, we also found that actions with an obvious motor incoordination (e.g., missed catch) also tended to cluster on the left side of the MDS space. Videos in this cluster consisted of strong motor incoordination (e.g., Video 5 featuring a dyadic interaction consisting of a series of missed fist bumps and handshakes). Meanwhile, the opposite (right side) of the horizontal dimension showed the cluster of actions with good motor coordination (e.g., Video 22 showing Donald Trump shaking Mitt Romney's hand) in the greeting behavior, suggesting that the degree to which the two actors coordinate their movements in the display can signal a key underlying dimension of awkwardness. Given the relationship between the horizontal dimension and the number of social words in the descriptions, we conjecture that motor coordination likely factors into participants' social judgments since motor coordination does not generally occur in isolated situations.

Touching duration may serve as a secondary cue for signaling awkward greeting behaviors. While cautious in our interpretation of the marginally significant relationship (Pritschet, Powell, Horne, 2016), this result is still consistent with previous findings that people are sensitive to temporal relation between actors (Sebanz & Knoblich, 2009; de la Rosa, et al., 2014; Burling & Lu, 2018) and different motion cues in actions (Thurman & Lu, 2016; Peng, Thurman & Lu, 2017). Furthermore, interpersonal touch also serves as a non-verbal social cue, incorporating important social information, such as emotional attributes and bonding (Gallace & Spence, 2010). Our results suggest that the correlations observed in the patch (most contextual information) and skeleton (least contextual information), but not in the body display (medium degree of contextual information) may be due the importance of touching duration as a social cue to awkwardness when incorporating key contextual information, such as the setting in which an interaction occurs (as in the patch display) and the importance of touching duration as a motor cue to awkwardness when relying predominantly on human kinematic information (as in the skeleton display). Further characterizing this relationship as it pertains to cultural differences, or remains inherent to American society, is an interesting area of future exploration.

As an important final point, while we aimed to separately examined the contribution of contextual information, consisting of rich social cues (e.g., scenery, attire), and human kinematic information (consisting of rich motor cues), as key signals underlying the evaluation of awkward interactions, the MDS results reveal their inextricable link. Specifically, the similar clustering of videos with both social and motor incoordination, as well as the key motor signals in touching duration (also related to social heuristics), prompt the following question: to what extent is motor coordination distinct from social coordination? Previous literature on interpersonal social interactions has shown that the temporal and motion congruency between two agents underlies human perceptions of social traits and/or animacy to the interaction (Thurman and Lu, 2014). Our present results similarly converge, suggesting that social coordination is likely impacted by motor coordination in awkwardness judgments. While objectively examining the extent of this relationship is outside the scope of our present paper, these results point to an important area of investigation that can even extend to wide-ranging, more ecologically valid domains, such as human robot interaction.

We conclude that the perception of awkwardness in greeting behaviors is based on general principles that significantly rely on motor cues, with the additional detection of failed social coordination for body movements that provide a key signal that a greeting has gone awry. Importantly, judging awkwardness does not appear to be entirely idiosyncratic, individuals appear to predominantly rely on a general set of heuristics rooted in human kinematics that is dynamically coupled with contextual information.

Acknowledgments

This study was supported by the National Science Foundation (BCS-1655300). We would like to thank Tianmin Shu with help with skeleton extraction, Tabitha Safari, Sophia Baia, Devin Bennett, and Jaime Wu for help with data collection.

References

Alter, A. L., Stern, C., Granot, Y., & Balcetis, E. (2016). The “bad is black” effect: Why people believe evildoers have darker skin than do-gooders. *Personality and Social Psychology Bulletin*, *42*(12), 1653-1665

Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of personality and social psychology*, *86*(4), 599.

Burling, J. M., & Lu, H. (2018). Categorizing coordination from the perception of joint actions. Attention, Perception, & Psychophysics, 80(1), 7-13.

Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2D pose estimation using part affinity fields. *The IEEE Conference on Computer Vision and Pattern Recognition*, 7291-7299.

de la Rosa, S., Choudhery, R. N., Curio, C., Ullman, S., Assif, L., & Bülthoff, H. H. (2014). Visual categorization of social interactions. Visual Cognition, 22(9), 1233–1271.

Freeman, J.B. & Ambady, N. (2011). A dynamic interactive theory of person construal. *Psychological Review*, *118*, 247-279.

Freeman, J.B., Penner, A.M., Saperstein, A., Scheutz, M., & Ambady, N. (2011). Looking the part: Social status cues shape race perception. *PLoS ONE*, *6*, e25107.

Freeman, J.B., Ma, Y., Han, S., & Ambady, N. (2013). Influences of culture and visual context on real-time social categorization. *Journal of Experimental Social Psychology, 49,* 206 210.

Heavey, L., Phillips, W., Baron-Cohen, S., & Rutter, M. (2000). The Awkward Moments Test: A naturalistic measure of social understanding in autism. *Journal of autism and developmental disorders*, *30*(3), 225-236.

Hohmann, T., Troje, N. F., Olmos, A., & Munert, J. (2011). The influence of motor expertise and motor experience on action and actor recognition. *Journal of Cognitive Psychology, 23*(4), 403-415.

Johansson, G. (1976). Spatio-temporal differentiation and integration in visual motion perception. *Psychological Research, 38*, 379-393.

Johnson, K. L., & Tassinary, L. G. (2005). Perceiving sex directly and indirectly: Meaning in motion and morphology. Psychological Science, 16(11), 890-897.

Kruskal, J. B., & Wish, M. (1978). Multidimensional scaling. Sage University paper series on quantitative applications in the social sciences, 07–011. Beverly Hills, CA: Sage.

Lin, G., Milan, A., Shen, C., & Reid, I. (2017). RefineNet: Multi-path refinement networks for high-resolution semantic segmentation. *The IEEE Conference on Computer Vision and Pattern Recognition*, 1925-1934.

Peng, Y., Thurman S., & Lu, H. (2017). Causal action: a fundamental constraint on perception and inference with body movements. Psychological Science, 28(6), 789-807.

Pollick, F. E., Paterson, H. M., Bruderline, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition, 82*(2), B51-B61.

Ren, X., Malik, J. (2003). Learning a classification model for segmentation. *Proceedings of the International Conference on Computer Vision*, 10–17.

Roether, C. L., Omlor, L., Christensen, A., & Giese, M. A. (2009). Critical features for the perception of emotion from gait. Journal of vision, 9(6), 15-15.

Runeson, S. & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: Expectation, gender recognition, and deceptive intention. *Journal of Experimental Psychology: General, 112*(4), 585-615.

Rule, N. O., & Ambady, N. (2008). Brief exposures: Male sexual orientation is accurately perceived at 50-ms. *Journal of Experimental Social Psychology, 44,* 1100-1105.

Rule, N. O., & Ambady, N. (2010). Democrats and Republicans can be differentiated from their faces. *PLoS ONE, 5,* e8733.

Rule, N. O., Ambady, N., Adams, R. B., Jr., & Macrae, C. N. (2008). Accuracy and awareness in the perception and categorization of male sexual orientation. *Journal of Personality and Social Psychology, 95,* 1019-1028.

Schmidt, R. C., Fitzpatrick, P., Caron, R., & Mergeche, J. (2011). Understanding social motor coordination. *Human movement science*, *30*(5), 834-845.

Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. Topics in Cognitive Science, 1(2), 353–367.

Snyder, M., Tanke, E. D., & Berscheid, E. (1977). Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology, 35*(9), 656-666.

Thurman, S., & Lu, H. (2014). Perception of social interactions for spatially scrambled biological motion. PLoS ONE. 9(11), 1-12.

Thurman, S. & Lu, H. (2016). Revisiting the importance of common body motion in human action perception. Attention, Perception & Psychophysics, 78(1), 30-36.

Todorov, A., Said, C. P., & Verosky, S. C. (2011). Personality impressions from facial appearance. In A. Calder, J. V. Haxby, M. Johnson, & G. Rhodes (Eds.), *Handbook of Face Perception* (pp. 631-652). Oxford University Press.

van Boxtel, J., & Lu, H. (2011). Visual search by action category. Journal of Vision. 11(7), 1-14.

van Boxtel, J., & Lu, H. (2012). Signature movements lead to efficient search for threatening actions. PLoS ONE , 7(5): e37085, 1-6. doi:10.1371/journal.pone.0037085.

van Boxtel, J. & Lu , H. (2013). A biological motion toolbox for reading, displaying and manipulating motion capture data in research settings. Journal of Vision, 13(12):7, 1-16.

van Boxtel, J. & Lu, H. (2015). Joints and their relations as critical features in action discrimination: Evidence from a classification image method. Journal of Vision. 15(1):20, 1-17.

CHAPTER 5

General Discussion

Across three studies, my dissertation aimed to examine the mechanisms that drive person recognition and processing from actions by revealing the visuomotor contributions underlying actions of ourselves and others. Between individuals, the aim was not only to measure the mechanistic distinctions that separate the self from others, but the commonalities at the action level that also help to construct the social concept of oneself.

In **Chapter 1**, we showed how self-recognition from our own actions, conveyed in visually impoverished point-light displays (PLDs), was modulated by the complexity of the action and intrinsic traits related to the participant. Self-recognition ability was importantly independent of lower-level actions factors related to speed or to perceived movement distinctiveness in the actions. This was an important finding, as it indicates that when motoric factors related to the action, or individual variance at the participant-level are accounted for, distinct visual cues to identity are less implicated in self-recognition. At first glance, this is somewhat in contrast to the results found by Coste et al (2021). Here, the researchers similarly captured participant body movements ($N = 15$), but had the participants performed improvised and simple postural motions (i.e., "keep knees extended with toes and heels in constant contact with the floor") that produced rhythmic motion patterns. The main finding from the study was that participants could recognize their own actions from postural signatures, but their recognition judgments of identity were largely based on their perceived self-similarity to the action. That is, the degree of visual similarity to oneself of the action they observed in the recognition task modulated their judgment as to whether the action belonged to oneself or another individual. These interpretations were fit in line with common coding theory (Prinz, 1997; Hommel, 2001), in that the closer the visual familiarity of the action matched one's own motor representation, the more likely the action would be attributed to the self.

Our findings do support, and further extend the results from Coste et al. In our study, we included a range of 27 actions and separated the actions into three categories based on their degree of motor planning. Across all action types that varied in motor planning, we found not only that participants could self-recognize their own actions, but substantial variability at the performance level for the action types. The action type (complex) that required the most motor planning was self-recognized significantly better than the action types that required less motor planning, which fits the premise of common coding theory, in that the increased motor familiarity resulting from more effortful motor planning facilitates the identity advantage for the self. It is important to note that with the increase in motor complexity from increased motor planning, the visual complexity of the action also increased. We used the Dynamic Time Warping (DTW) algorithm to compute the movement dissimilarity (or distinctiveness) across individuals in the self-recognition task. More distinctive visual movements were evident in actions required the most motor planning (complex). Notably, however, actions that required the least motor planning based on copying another agent's sequenced motor repertoire (imitation), also conveyed a very high degree of movement distinctiveness due to the range of variable actions that were included in the action category. As a result, movement distinctiveness was not significantly different for imitation actions and complex actions, contradictory to self-recognition performance. From this pattern of results, it is clear holistically that movement distinctiveness of the actions cannot be the driving factor in self-action recognition. Familiarity with one's own motor repertoire (i.e., motor planning of the participant) appeared to be the main contributor to self-recognition that we attributed to motor experience. To confirm this, we conducted multiple regression analyses including correlational and linear mixed effects modeling, which further confirmed the importance of motor experience at the action level to self-recognition.

Returning to Coste et al (2021), it appears that movement distinctiveness likely plays a central role to self-recognition when other factors related to the action are not accounted for. However, by accounting for differences in motor planning and the complexity of the actions, the contribution of movement distinctness was negligible, as was found in our results. Our methodology to compute movement distinctiveness using DTW did have some limitations as discussed in the results for Chapter 1. It is important that future studies augment this work –perhaps using finer-grained measures of movement distinctiveness, or perhaps directly manipulating the degree of visual similarity of the PLD to oneself, by introducing some level of deviance bias (e.g., angular deviations), typically studied in paradigms measuring agency of actions (e.g., Penton et al., 2022).

Beyond measuring motor familiarity in isolation, our study found interactions with the motoric factors and intrinsic traits of the participants linked to Autism, Schizotypy, and motor imagery. Disturbances in self-processing have been reported across the Schizophrenia-spectrum and Autism-spectrum. In Schizophrenia, self-disturbances (see Nelson, Whitford, Lavoie, & Sass, 2014 for a review) are thought to account for the hallucinatory experiences prevalent in the condition, as a form of misattributing self-generated stimuli to external sources. Further, numerous studies have reported atypical self-reflective processing (Van der Meer, Costafreda, Aleman, David, 2010), body ownership (e.g., Thakkar, Nichols, McIntosh, Park, 2011; Klaver and Dijkerman, 2016), self-monitoring (Keefe, Arnold, Bayen, McEvoy, & Wilson, 2002; Carter et al., 2001), agency (e.g., Jeannerod, 2009), and self-control (Frith, 2005) in Schizophrenia, as well as converging evidence in ASD, showing developmental delays in self-face recognition ability (Spiker & Ricks, 1984; Dawson & McKissick, 1984).

While the prevailing empirical method to study self-recognition of whole-body movements is to significantly degrade visual cues to identity, the impact of intrinsic traits, and in relation to these clinical conditions, to recognition ability has gained little attention within the literature. There are a number of reasons as to why it is important to measure individual difference traits in self-action recognition. First, the unique contribution of various individual difference measures can uncover critical information that could potentially be lost through group-level averaging (Peterzell & Kennedy, 2016). Second, self-recognition is a complex process, with its investigation particularly hampered by its own operationalization and resulting lack of objectivity (consisting of no clear-cut computational investigation). Finally, examining individual difference traits linked to clinical disorders (i.e., schizophrenia and autism) in the general population has several advantages over utilizing clinical patients, notably in its cost-effectiveness and lack of confounding medication that could otherwise hamper with the interpretation of results (Lam, Raine, and Lee; 2017).

Several findings emerged from the individual difference results. Notably, we found interactions between the motoric factor and variability for subscales of the composite individual difference measure. Subscale variability ranged from both motoric dimensions to social and interpersonal dimensions, that impacted and moderated (motor imagery and autism-spectrum) the relationship to self-recognition. These results are indicative as to how the automaticity of self-recognition can be misleading. Rather than a simple process, self-action recognition involves a complex architecture with dynamic interactions between perceptual and motor systems, as well as intrinsic variance related to the participant.

In **Chapter 2,** we extended the behavioral work on self-recognition to brain imaging. Our aim in the study was based on the premise of common coding theory and the pattern of results related to motor experience/expertise from Chapter 1—specifically whether common codes as reflected by the mirror neuron frontoparietal regions would be more strongly implicated in self-action recognition relative to others. We further integrated the work with the visual neuroscience paradigms using PLDs of the whole-body, as our stimulus set was purposely different from typical those typically used cognitive neuroscience action processing tasks that tend to involve isolated body parts or social neuroscience tasks on high-level attributes of the self. Moreover, by characterizing a broader range of regions within the action observation network (AON), our findings integrated regions commonly targeted in the visual neuroscience tasks involving the temporooccipital cortex including the extrastriate body area (EBA) and posterior superior temporal sulcus (pSTS), as well as regions found in cognitive neuroscience such as the IPL and IFG.

Our findings are indicative of a specialized pathway for self-recognition that showed increased recruitment of the AON. We found activity across regions in the AON comprising both temporooccipital and frontoparietal regions. Specifically, temporooccipital regions within the AON were involved in domain-general identity decoding, while frontoparietal regions within the AON were broadly attuned to self-recognition. These results appeal to how the intrinsic nature of self-recognition relies on sensorimotor properties. Moreover, our findings also reveal the common connection for self and others by demonstrating how processing for both, recruits common neural circuitry across the AON.

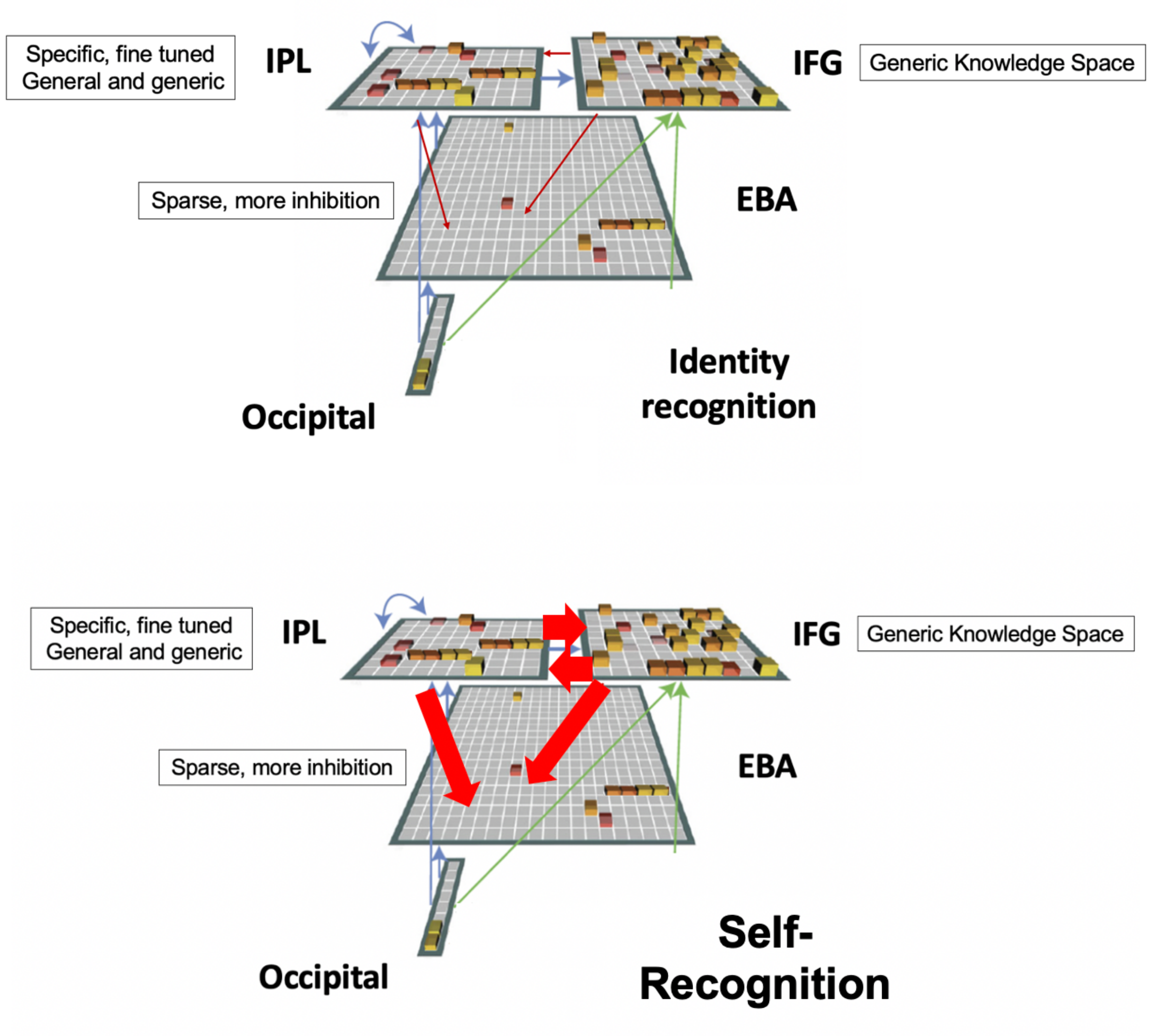


Figure 31. *Top:* Possible network architecture for general action recognition from whole-body movements in PLDs from **Chapter 2**. Red arrows: feedback influence from IFG and IPL. Blue arrows: feedforward influence from EBA to higher-level. Green: reflect idealized feedforward processing from sensory (occipital) to evaluative processing (IFG). *Bottom:* Greater top-down influence for whole-body self-action recognition reflected through feedback influence from the IFG and IPL (red arrow). Modeled using Emergent (Aisa et al., 2008), Leabra (O'Reilly, 1996) and altering the model architecture from Schapiro et al (2017).

Considering the increased recruitment of the AON found for self-recognition, one possible framework for self-recognition could involve decoding attuned to specific features at the temporooccipital regions (in this case localized to the EBA), as depicted in *Figure 31*. Here, I focus on three regions: EBA, IPL, and IFG based on the results from the gPPI analysis in Chapter 2, as well as previous effective connectivity work (e.g., Sasaki et al., 2018). Since our functional connectivity analysis was nondirectional, *Figure 31* represents a potential framework for the directionality of information flow, as well as their representational structure (based on the results from the RDA analyses). Input is presumed to arrive from the occipital regions which then is integrated in the EBA. Here, the EBA has sparse and non-overlapping representations, allowing for granular discrimination of postural and kinematic signatures as cues to identity. Information from the EBA flows upward to the IPL, which contains populations of neurons that encode both specific and fine-tuned features, as well as generic and generalizable populations that encode identity. Hence, both action features (more anteriorly) and identity features (more posteriorly) are represented here. Information from the IPL could then be passed on the IFG, which encodes information at an evaluative and generic knowledge space, with high-density and overlapping representations. This may explain why we identified cues to movement distinctiveness represented here, which importantly did not overlap with identity. Finally, the entire model hierarchy involves directed feedback loops from the IFG to the temporooccipital regions (EBA/pSTS) and IPL for more evaluative judgments. For self-identification (Figure 31, bottom), the architecture infers greatest top-down influence from the IFG and IPL relative to identification of others.

In Chapter 3, we moved away from contextually impoverished PLDs and incorporated naturalistic video stimuli of social interactions. We went in the opposite direction methodologically, where we started with a high degree of context in one level of the independent variable, and parametrically degraded the amount of context in the display to comprise the other two levels. The goal of this study was to assess the impact of social context and the measure the underlying visuomotor features that contribute to social attribution judgments of interpersonal congruency. Specifically, we asked whether systematizing the degree of context affects judgments of action congruency— "awkwardness"— of social interactions of others. Similar to Chapters 1 and 2, we aimed to explore the visuomotor space of the features that contribute to interactions of others. The interactions entailed both regular (natural) interactions, as well as those that violated our natural sense of congruency, termed "awkward." We constructed the database of stimuli as compiled from online resources and then performed a stimulus validation check in experiment 3.1, where participants categorized each interaction as natural or awkward. In 3.2, we used advances in computer vision and image processing to systematically degrade the degree of context in the interactions into three types of stimuli: (1) patch, that blurred the interaction scene but preserved the highest degree of social context (2) body, that eliminated most identity features and background, except for the body morphology and structure (3) skeleton, that only preserved the body structure. Importantly, kinematics of the actions were held constant across all three displays, which allowed us to explore how featural representations related to the body and visual scene affected participant judgments of incongruency/awkwardness of the interaction.

Across the three displays, we found reliable consistency across participants in categorizing interactions as awkward or natural. Importantly, even after degrading the social context to just kinematics conveyed by stick figures, participants were still consistent in terms of extracting high-level social information from the displays (i.e., determining whether the interaction was awkward or not). This corroborates work from Chapters 1 and 2, where we show how participants can extract high-level attributes, such as identity, from visually impoverished stimuli, and broadens the results to judgments within naturalistic social interactions. In line with the importance of the motor system to action processing, these findings suggest high-level social information can be grasped even on the basis of motoric cues (e.g., kinematics), which at the neural level, could potentially be facilitated by frontoparietal mirroring systems in the brain.

We further explored the representational space of the social judgments related to the interactions that showed a high degree of incongruency. Two factors accounted for some of the explanatory variance in the awkwardness judgments. The featural space of the awkwardness judgments represented coordination related to both interpersonal social (number of social words), as well as motor (touching duration) factors. The motoric factor contribution was marginally significant, but only in the skeleton display. Hence, while social context played a relatively a strong role in influencing judgments of awkwardness, degrading the social context increased the involvement that motoric cues play, even in social attributions. Evidenced even at the behavioral level, we show how social judgments are largely inferred by both social and motoric factors.

There are limitations in Chapter 3 that should be addressed in future work. First, explanatory dimensions related to other social and visuomotor features should be measured. We did not sufficiently explain all the variability in the 2-D space for the rating judgments for the display types. Part of the reason is due to the fact that we did not include a wide range of tests related to motoric features (e.g., speed, acceleration). Moreover, social features are hard to quantitatively measure given their latent structure. Potentially incorporating additional methods to quantitatively measure social information in the display types, such as Word2Vec or classic methods in natural language processing could allow for finer-grained subtypes of social relations to emerge in the representational space. Additionally, while it is certainly beneficial for empirical paradigms to use naturalistic stimuli, they also present some issues of experimental control. Our paradigm included a range of videos that varied in social context and types of interactions, as means to mitigate the influence of any single factor influencing the social judgments. However tighter control, while still incorporating naturalistic types of paradigms, would be beneficial in future research.

Conclusion

In general summary, the dissertation sheds light on the important contribution of the visuomotor mechanisms to high-level attributes related to actions of the self and others. Featurally, motor cues appear to be a common contributing link, whether presented in an isolated self-recognition task, or in naturalistic interactions that involve multiple individuals. Yet, as is evident in neuropsychology, the research emphasis remains largely constrained to studying higher-level cognitive processes, rather than on the causal motoric factors that underlie them. Producing movements and forming goal-directed actions are one of the most common experiences we have. By bridging action processing methods in visual and cognitive neuroscience, my dissertation takes steps to attribute these visuomotor mechanisms to high-level social processes, and ultimately in how they relate to our own self-representation.

References

Aisa, B., Mingus, B., & O'Reilly, R. (2008). The emergent neural modeling system. *Neural networks*, *21*(8), 1146-1152.

Coste, A., Bardy, B. G., Janaqi, S., Słowiński, P., Tsaneva-Atanasova, K., Goupil, J. L., & Marin, L. (2021). Decoding identity from motion: how motor similarities colour our perception of self and others. Psychological research, 85(2), 509-519.

Dawson, G., & McKissick, F. C. (1984). Self-recognition in autistic children. Journal of Autism and Developmental Disorders, 14(4), 383-394.

Frith, C. (2005). The self in action: Lessons from delusions of control. Consciousness and cognition, 14(4), 752-770.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. Behavioral and brain sciences, 24(5), 849-878.

Jeannerod, M. (2009). The sense of agency and its disturbances in schizophrenia: a reappraisal. Experimental Brain Research, 192(3), 527-532.

Keefe, R. S., Arnold, M. C., Bayen, U. J., McEvoy, J. P., & Wilson, W. H. (2002). Source-monitoring deficits for self-generated stimuli in schizophrenia: multinomial modeling of data from three sources. Schizophrenia research, 57(1), 51-67.

Klaver, M., & Dijkerman, H. C. (2016). Bodily experience in schizophrenia: factors underlying a disturbed sense of body ownership. Frontiers in human neuroscience, 10, 305.

Nelson, B., Whitford, T. J., Lavoie, S., & Sass, L. A. (2014). What are the neurocognitive correlates of basic self-disturbance in schizophrenia?: Integrating phenomenology and neurocognition. Part 1 (Source monitoring deficits). Schizophrenia research, 152(1), 12-19.

O'Reilly, R. C. (1996). The Leabra model of neural interactions and learning in the neocortex

Penton, T., Wang, X., Catmur, C., & Bird, G. (2022). Investigating the sense of agency and its relation to subclinical traits using a novel task. *Experimental brain research*, *240*(5), 1399-1410.

Peterzell, D. H., & Kennedy, J. F. (2016). Discovering sensory processes using individual differences: A review and factor analytic manifesto. Electronic Imaging, 2016(16), 1-11.

Rosenbaum, D. A., & Feghhi, I. (2019). The time for action is at hand. *Attention, Perception, & Psychophysics*, *81*(7), 2123-2138.

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.

Spiker, D., & Ricks, M. (1984). Visual self-recognition in autistic children: Developmental relationships. Child development, 214-225.

van der Meer, L., Costafreda, S., Aleman, A., & David, A. S. (2010). Self-reflection and the brain: a theoretical review and meta-analysis of neuroimaging studies with implications for schizophrenia. Neuroscience & Biobehavioral Reviews, 34(6), 935-946.

A1. Supplementary Materials Chapter II

I. DESCRIPTIVE STATISTICS

Measure	*N*	Max Score	Mean	Standard Deviation	Range
AQ	98	50	18.62	6.03	8-38
AQ-Comm					
AQ-SocialSkill					
AQ-Imagination	98	10	2.49	1.93	0-8
AQ-	98	10	2.49	2.21	0-10
AttentionDetail	98	10	2.67	1.45	0-7
AQ-	98	10	5.36	2.18	0-10
AttentionSwitch	98	10	5.60	1.95	1-9
SPQ	98	74	23.54	11.14	3-56
SPQ-	98	9	2.56	1.95	0-9
UnusualPerceptual	98	9	3.81	2.39	0-9
SPQ-	98	7	1.15	1.39	0-5
IdeasReference	98	8	2.66	1.96	0-8
SPQ-OddBeliefs	98	8	4.09	2.33	0-8
SPQ-Suspiciousness					
SPQ-SocialAnxiety					
SPQ-NoCloseFriends	98	9	2.29	2.20	0-9
SPQ-OddBehavior	98	7	1.15	1.99	0-7
SPQ-OddSpeech	98	9	3.39	1.93	0-9

SPQ-	98	8	1.75	1.74	0-7
ConstrictedAffect	98	180	67.95	20.16	36 – 157
VMIQ-2	98	60	26.13	10.19	12 – 49
VMIQ-E	98	60	20.70	7.32	12 – 49
VMIQ-I	98	60	21.11	8.64	12 – 60
VMIQ-K					

Table 1. Descriptive statistics for all composite (AQ, SPQ, VMIQ-2) and subscale measures

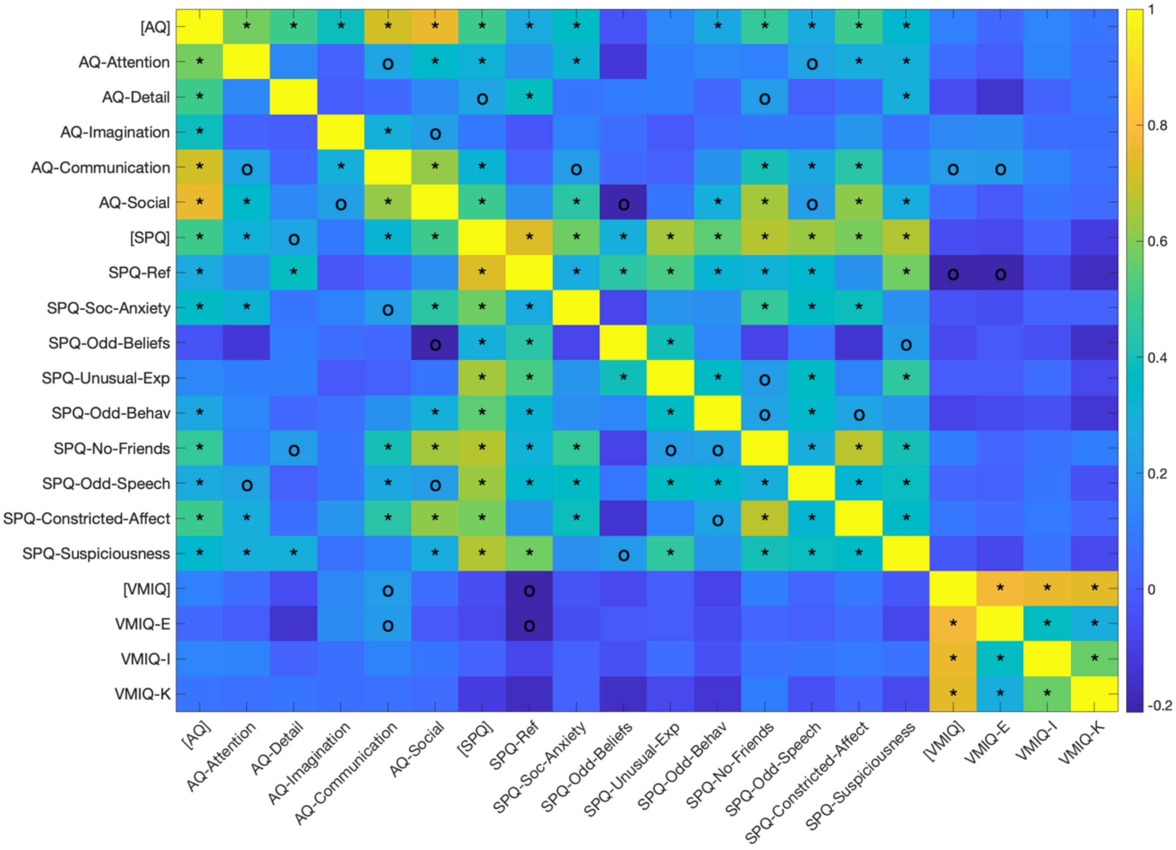
Independent sample t-tests revealed no significant effects of gender on composite AQ ($\boldsymbol{t}(96) = 1.06$, $\boldsymbol{p} = .293$, $\boldsymbol{d} = .239$), SPQ ($\boldsymbol{t}(96) = 0.01$, $\boldsymbol{p} = .994$, $\boldsymbol{d} = .002$), or VMIQ-2 ($\boldsymbol{t}(96) = -0.98$, $\boldsymbol{p} = .328$, $\boldsymbol{d} = -.222$) scores. Table 1 include descriptive statistics for all composite and subscale scores in individual measures. Figure S1 shows correlations between all the measures.

II. CONFIDENCE DATA

Metacognitive assessments of self-recognition performance

Seventy-four participants provided confidence ratings for the self-recognition judgments. One-way repeated measures ANOVA revealed participants' confidence ratings for self-recognition reflected performance accuracy based on action type, $\boldsymbol{F}(2,146) = 6.35$, $\boldsymbol{p} = .002$, $\eta_p^2 = .080$. Participants were significantly more confident for self-recognition for complex than for simple actions ($\boldsymbol{t}(73) = 4.05$, $\boldsymbol{p} < .001$, $\boldsymbol{d} = .471$), and for imitation actions ($\boldsymbol{t}(73) = 2.04$, $\boldsymbol{p} = .046$, $\boldsymbol{d} = .237$), while there was no difference in confidence judgments between imitation and simple actions ($\boldsymbol{t}(73) = 1.25$, $\boldsymbol{p} = 0.217$, $\boldsymbol{d} = .145$).

III. INTRINSIC MEASURES



A1. Spearman correlation heatmap depicts spearman rank-order relationships between all intrinsic measures (composite scores in brackets). Brighter colors indicate stronger relationships between trait measures. Circles denote $.01 < \boldsymbol{p} < .05$; asterisks denote $\boldsymbol{p} < .01$ significance.

IV. SELF-RECOGNITION VS VISUAL RECOGNITION

To confirm the importance of motor experience to action identity recognition (speculated upon in Loula et al., 2005), we included an additional visual recognition task in which participants were asked to recognize previously-seen and imitated actions. We expected that the impact of visual experience on identity recognition may diminish after a long delay of about one month, but motor experience of own body movements would maintain in its effect on self-action recognition.

The first 28 participants only underwent the self-recognition task. The subsequent 73 participants were tested with both the self-recognition task and an additional visual recognition task consisting of nine trials depicting the forward-facing imitation actions. In the visual recognition task, participants were instructed to identify the stick-figure actor they imitated during the motion recording phase (converted to point-light display) amongst three other distractor point-light actors who performed the same action, that were identical to those used in the self-recognition task. The visual recognition task allowed us to compare performance of recognizing self-actions versus recognition of imitated actions performed by others, which delineates self-generated from visually observed actions. The visual recognition trials used the identical stimulus layout as in the self-recognition task, except that the participants' own action in the self-recognition trials was replaced by the point-light actions of original imitation actor from the CMU motion capture database. The order of presentation of the visual recognition task was counterbalanced to either follow or precede the self-recognition task across participants.

We contrasted performance in the self-recognition task in comparison to performance in the visual recognition task for imitation actions. Self-generated actions present the unique case of expert motor experience (kinesthetic, proprioceptive, and muscle contraction information), but impoverished visual experience from a third-person perspective. By contrast, expertise of movement patterns in many situations, such as sports, show both superior motor ability in performance of one's own actions, but also superior visual ability in recognition of actions performed by others, such as teammates (Hohmann et al., 2011). Our task manipulation allowed for the comparison between actions that participants performed (i.e., own imitation action) versus actions that the participant observed (i.e., stick figure imitation actor). As shown in Figure S1, participants performed at chance level for visual recognition of previously viewed actions ($\boldsymbol{M} = 0.23$, $\boldsymbol{SD} = 0.19$), $\boldsymbol{t}(72) = -\ 0.718$, $\boldsymbol{p} = .475$, $\boldsymbol{d} = -.084$, but showed much greater recognition performance in identification of one's own action ($\boldsymbol{M} = 0.38$, $\boldsymbol{SD} = 0.15$), with visual recognition performance significantly lower than performance in self-identification, $\boldsymbol{t}(72) = 5.41$, $\boldsymbol{p} < .001$, $\boldsymbol{d} = .633$. The around-chance performance for actions with only visual experience confirms that prior visual experience alone may not suffice for supporting self-recognition. Identifying oneself relies on the ability to simulate the action onto one's own motor system, with visual self-recognition in turn dependent on a matching process to compare simulated action to observed action.

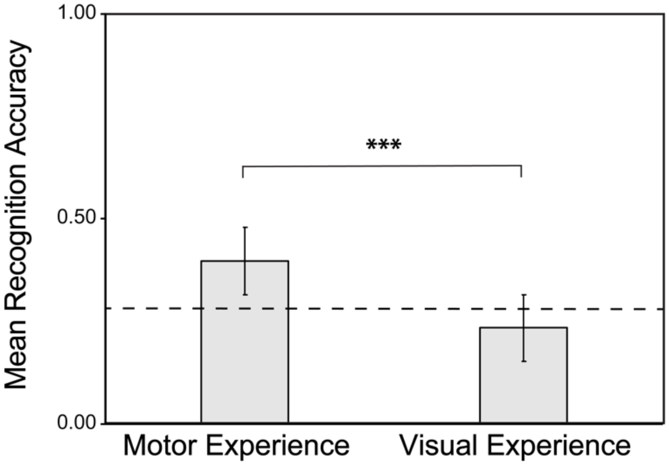


Figure A2. Recognition accuracy for imitation actions in the self-recognition task with motor experience versus performance in the visual-recognition task with visual experience. Significantly weaker performance from recognizing actions from visual experience than for self-recognition from performed actions. Dashed line indicates chance performance (0.25). Error bars indicate standard error of means.

V. SPEED ANALYSIS

To measure the contribution of movement speed to self-recognition, we calculated a movement distinctiveness value for every participants' action within the three action types (simple, complex, imitation) and measured the relation to self-recognition performance for the action types. For each action, we computed the average 3D positional displacement across all frames and all 17 joints relative to all other participants as a measure of speed distinctiveness. No significant differences in speed distinctiveness were found between the action types: complex v simple**,** $\boldsymbol{t}(100) < .001$, $\boldsymbol{p} = 1.00$, cohen's $\boldsymbol{d} = .009$; complex v imitation: $\boldsymbol{t}(100) < .001$, $\boldsymbol{p} = 1.00$, cohen's $\boldsymbol{d} = .011$; simple v imitation: $\boldsymbol{t}(100) < .001$, $\boldsymbol{p} = 1.00$, cohen's $\boldsymbol{d} = .009$), in contrast to the performance pattern for self-recognition. Addiitonally, no correlation was found between speed distinctiveness and accuracy for the action types: simple (spearman $\boldsymbol{\varrho} = .332$, $\mathrm{p} = .059$), complex (spearman $\boldsymbol{\varrho} = .151$, $\boldsymbol{p} =.401$), imitation (spearman $\boldsymbol{\varrho} = .058$, $\boldsymbol{p} = .747$).

A2. Supplementary Materials Chapter III

I. Univariate table of results for subtraction contrasts of interest (self vs others), using FSL randomise, corrected using TFCE ($p < .05$)

Peak locations for self contrasts

Contrast	Hemisphere	Area	*Peak MNI (x,y,z)*	Max Z	Cluster Size	*P*
Self > Stranger	Left	Supramarginal Gyrus	(-62, -48, 28)	3.72	463	.0181
	Left	*Sub-cluster:* Angular Gyrus	(-60, -52, 30)			
	Left	Insula	(-42, 10, -8)	3.72	160	.0228
	Left	Anterior Cingulate Cortex	(-2, 20, 18)	3.72	55	.040
	Right	Anterior Cingulate Cortex	(4, 14, 28)	3.72	19	.040
Self > Friend	Left	Supramarginal gyrus	(-54, -50, 30)	3.72	378	.036

II. BEHAVIORAL RESULTS FOR ACTION TYPE

A 2 x 3 repeated measures ANOVA revealed a significant interaction between action type (verbal vs visual instruction) and identity (self, friend, and stranger) ($F(2,38) = 7.79$, $p = .001$, partial eta square = .291). We found that action type only modulated the performance difference for self-generated actions. Actions that were freely performed (i.e., verbal instruction; $M = .616$, $SD = .199$) were recognized at greater rates than for actions that were imitated (i.e., visual instruction; $M = .509$, $SD = .188$), $t(19) = 3.32$, $p = .004$, d = .743. Greatest accuracy for freely performed actions only held true for self-generated actions. By contrast, a marginal difference was observed (after bonferonni correction; $\text{alpha}_{\text{new}} = .016$) in the reverse direction for recognition of stranger actions, such that participants showed marginally better recognition for actions that were more stereotyped (i.e., visually instructed) than verbally instructed ($t(19) = 2.178$, $p = .042$, $d = .487$).

III. RDM FEATURAL CALCULATIONS

Movement distinctiveness:

The following steps were implemented for Dynamic Time Warping (DTW) analyses in MATLAB R2020a:

(1) For each participant's actions, we extracted the 3D positions of each of the 25 joints using the BioMotion toolbox (van Boxtel & Lu, 2013).

(2) We centered each trajectory of a joint to zero in order to remove the impact of global factors (e.g., global body displacements, limb length, etc.) on the similarity measures.

(3) We then implemented an action DTW algorithm (Pham, Le, & Le, 2014) to search for a temporal warping function shared across all 25 joints.

(4) After deriving the optimal warping function, the analysis computes the frame-by-frame Euclidean distances of the temporally warped joint trajectories in actions performed by different actors.

(5) DTW distance was then computed as the sum of the distances between all joint trajectories normalized by the number of frames of a target actor. This normalization step is required in order to account for the different durations across participants performing the same action

(6) For each participant, the dissimilarity of the target participant performing an action from all other identities was captured by a mean DTW distance measure, computed by averaging across pairwise DTW distances between the target participant with the other actors (friend, stranger) in performing this action to construct the 36 x 36 representational dissimilarity matrix (RDM).

Speed and Acceleration Distinctiveness:

To measure the contribution of movement speed to self-recognition, we calculated a movement distinctiveness value for every participants' action for each individual action in MATLAB R2020a. For each action, we computed the average 3D positional displacement across all frames and all 25 joints (using the first-order derivative of position) extracted from Biomotion Toolbox (Van Boxtel & Lu, 2013). We then computed the average pairwise Euclidean distance to all other identities as a measure of speed distinctiveness to construct the 36 x 36 RDM. No relationships were found between speed distinctiveness and self-recognition performance (spearman ρ = -.077, p =.747), nor with general identity discrimination (spearman ρ = -.093, p =.698).

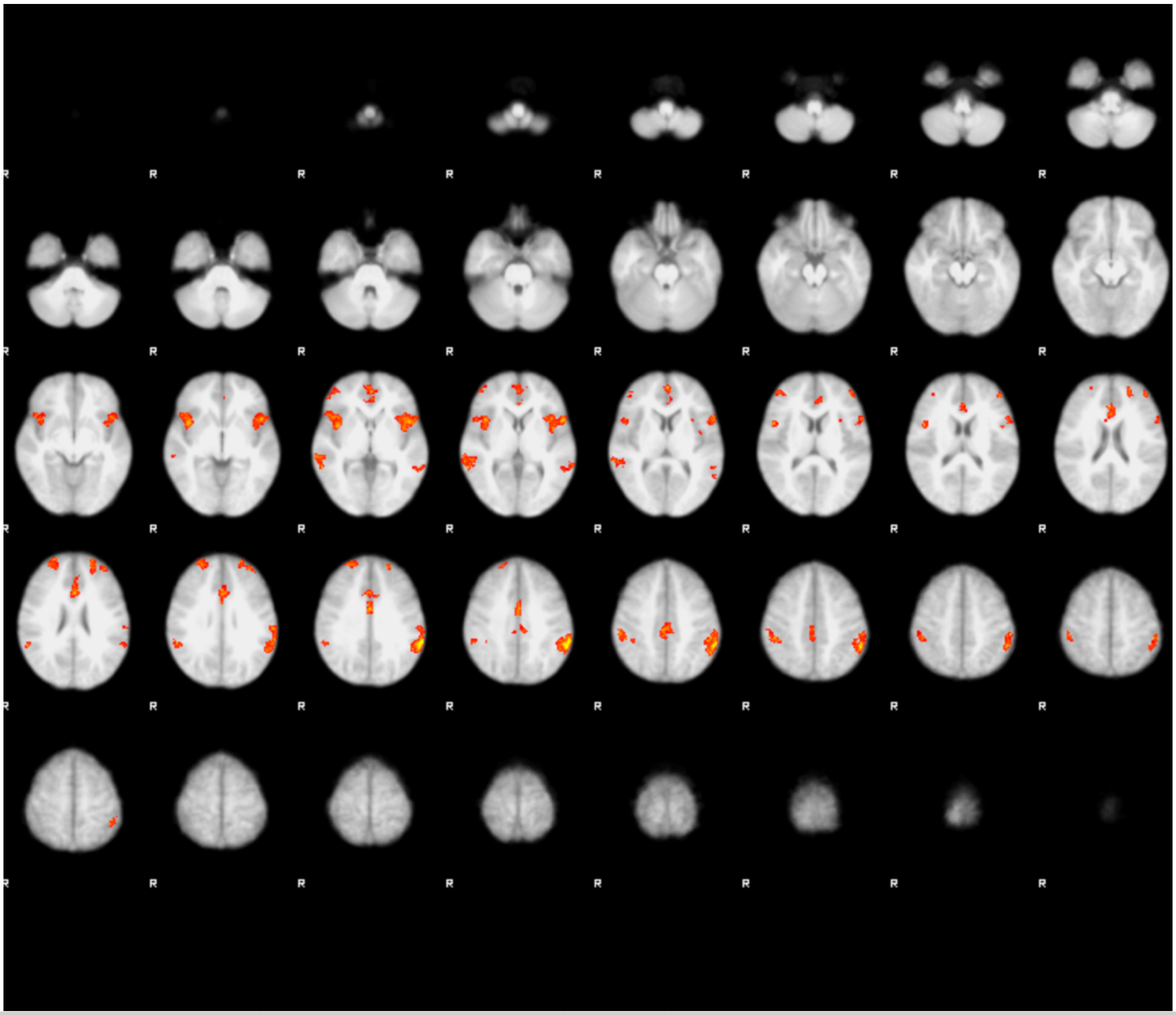
Acceleration was computed in the identical way, but instead taking the second-order derivative of position. No relationships were found between acceleration distinctiveness and self-recognition performance (spearman $\rho = -.234$, $p =.321$), nor with general identity discrimination (spearman $\rho = -.272$, $p =.247$).

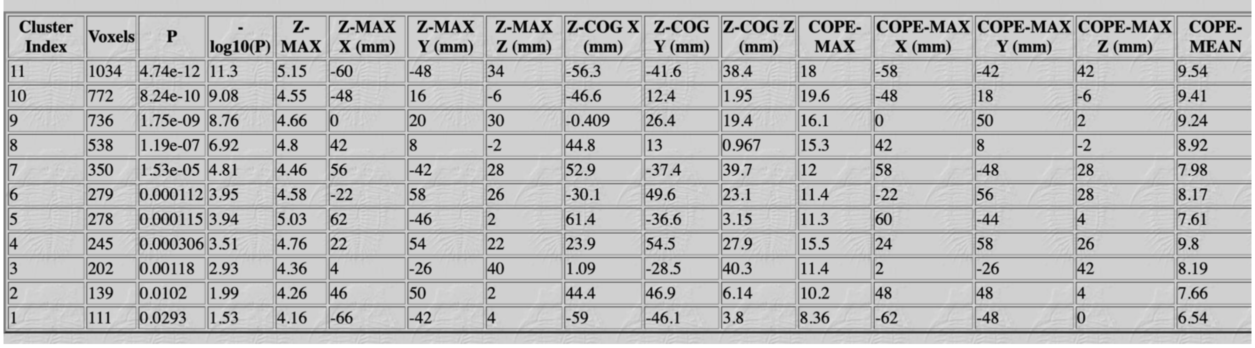
Postural Distinctiveness:

To compute body structure distinctiveness, we measured the averaged postural limb length for each participant. We used Biomotion Toolbox (Van Boxtel & Lu, 2013) to extract the 25 normalized joint coordinates for each participant and compute their relative Euclidean distance, which resulted in 24 limbs produced for each actor. We then computed the average pairwise Euclidean distance across all limb lengths between each of the identities (self, friend, stranger) to construct the 36 x 36 RDM for each participant. No relationships were found between body structure distinctiveness and self-recognition (spearman $\rho = -.195$, $p =.410$), nor with general identity discrimination (spearman $\rho = .053$, $p =.823$).

IV. Univariate Cluster Correction (as reported by FSL)

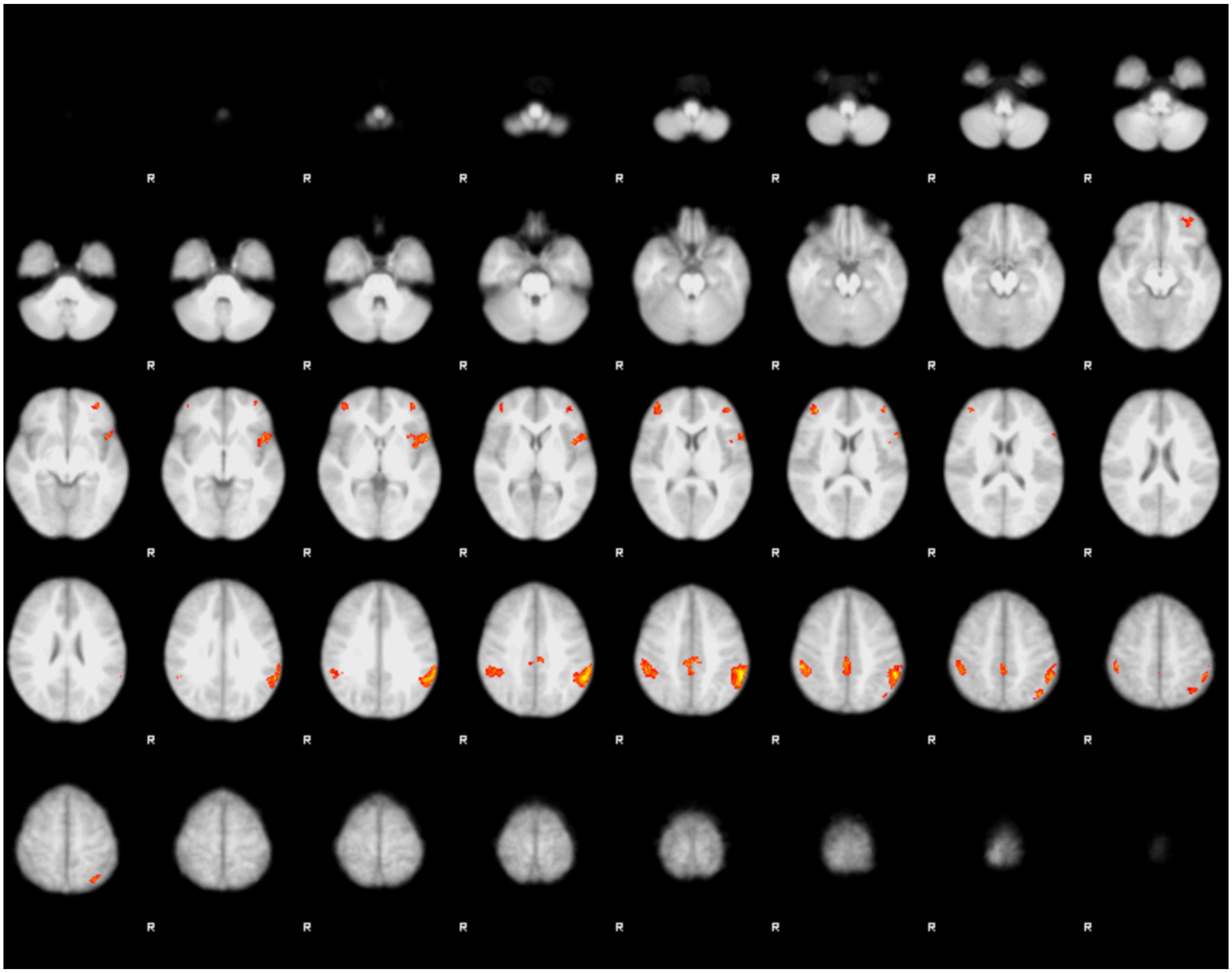
Self > stranger ($Z > 3.1$, $p < .05$; RFT Cluster Correction):

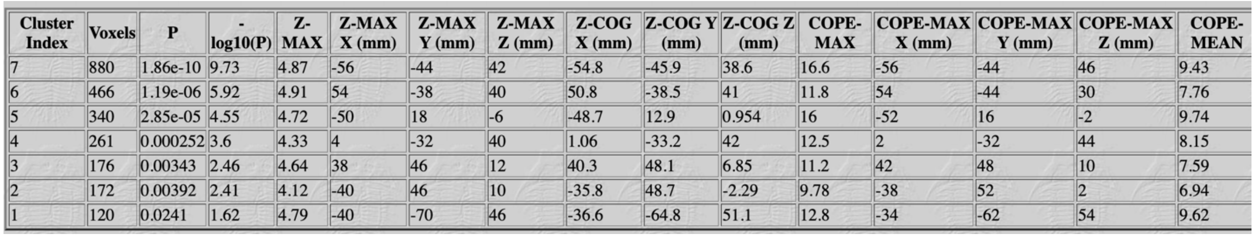




Cluster Index	Voxels	P	-log10(P)	Z-MAX	Z-MAX X (mm)	Z-MAX Y (mm)	Z-MAX Z (mm)	Z-COG X (mm)	Z-COG Y (mm)	Z-COG Z (mm)	COPE-MAX	COPE-MAX X (mm)	COPE-MAX Y (mm)	COPE-MAX Z (mm)	COPE-MEAN
11	1034	4.74e-12	11.3	5.15	-60	-48	34	-56.3	-41.6	38.4	18	-58	-42	42	9.54
10	772	8.24e-10	9.08	4.55	-48	16	-6	-46.6	12.4	1.95	19.6	-48	18	-6	9.41
9	736	1.75e-09	8.76	4.66	0	20	30	-0.409	26.4	19.4	16.1	0	50	2	9.24
8	538	1.19e-07	6.92	4.8	42	8	-2	44.8	13	0.967	15.3	42	8	-2	8.92
7	350	1.53e-05	4.81	4.46	56	-42	28	52.9	-37.4	39.7	12	58	-48	28	7.98
6	279	0.000112	3.95	4.58	-22	58	26	-30.1	49.6	23.1	11.4	-22	56	28	8.17
5	278	0.000115	3.94	5.03	62	-46	2	61.4	-36.6	3.15	11.3	60	-44	4	7.61
4	245	0.000306	3.51	4.76	22	54	22	23.9	54.5	27.9	15.5	24	58	26	9.8
3	202	0.00118	2.93	4.36	4	-26	40	1.09	-28.5	40.3	11.4	2	-26	42	8.19
2	139	0.0102	1.99	4.26	46	50	2	44.4	46.9	6.14	10.2	48	48	4	7.66
1	111	0.0293	1.53	4.16	-66	-42	4	-59	-46.1	3.8	8.36	-62	-48	0	6.54

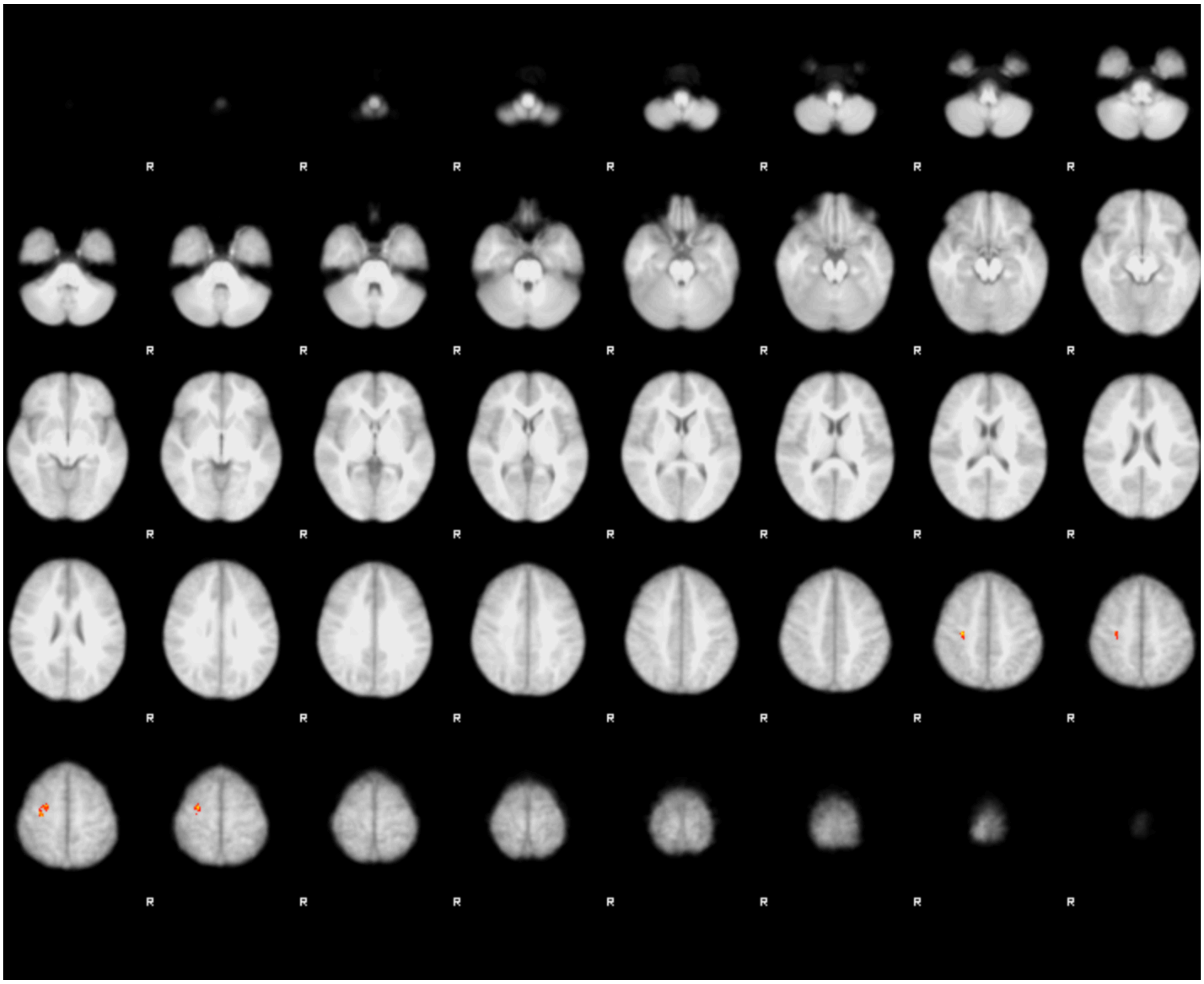
Self > Friend ($Z > 3.1$, $p < .05$; RFT Cluster Correction):

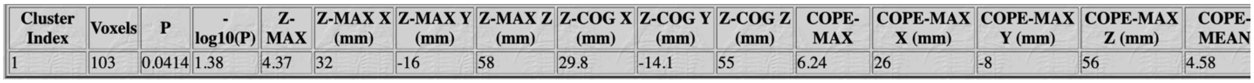




Cluster Index	Voxels	P	-log10(P)	Z-MAX	Z-MAX X (mm)	Z-MAX Y (mm)	Z-MAX Z (mm)	Z-COG X (mm)	Z-COG Y (mm)	Z-COG Z (mm)	COPE-MAX	COPE-MAX X (mm)	COPE-MAX Y (mm)	COPE-MAX Z (mm)	COPE-MEAN
7	880	1.86e-10	9.73	4.87	-56	-44	42	-54.8	-45.9	38.6	16.6	-56	-44	46	9.43
6	466	1.19e-06	5.92	4.91	54	-38	40	50.8	-38.5	41	11.8	54	-44	30	7.76
5	340	2.85e-05	4.55	4.72	-50	18	-6	-48.7	12.9	0.954	16	-52	16	-2	9.74
4	261	0.000252	3.6	4.33	4	-32	40	1.06	-33.2	42	12.5	2	-32	44	8.15
3	176	0.00343	2.46	4.64	38	46	12	40.3	48.1	6.85	11.2	42	48	10	7.59
2	172	0.00392	2.41	4.12	-40	46	10	-35.8	48.7	-2.29	9.78	-38	52	2	6.94
1	120	0.0241	1.62	4.79	-40	-70	46	-36.6	-64.8	51.1	12.8	-34	-62	54	9.62

Stranger > Self: (Z > 3.1, p < .05; RFT Cluster Correction):





Cluster Index	Voxels	P	-log10(P)	Z-MAX	Z-MAX X (mm)	Z-MAX Y (mm)	Z-MAX Z (mm)	Z-COG X (mm)	Z-COG Y (mm)	Z-COG Z (mm)	COPE-MAX	COPE-MAX X (mm)	COPE-MAX Y (mm)	COPE-MAX Z (mm)	COPE-MEAN
1	103	0.0414	1.38	4.37	32	-16	58	29.8	-14.1	55	6.24	26	-8	56	4.58

No significant clusters obtained for friend > self; friend > stranger (Z > 3.1, p < .05)

A3. Supplementary Materials Chapter IV

Links to Video Stimuli

Clip name	Link	Start	Duration (s)
clip1	https://youtu.be/T84se4fc4KU?t=30s	0:31	6.72
clip2	https://youtu.be/_npyQonU-V0?t=1m39s **(N/A)**	4:72	3.80
clip3	https://www.youtube.com/watch?v=i6i1PpEtM_4	1:06	10.04
clip4	https://youtu.be/6yWCmMQy1Qk?t=1m45s	1:44	3.80
clip5	https://www.youtube.com/watch?v=c5aN5zOEpM8	0:05	26.96
clip6	https://youtu.be/T84se4fc4KU?t=28s	0:27	3.08
clip7	https://www.youtube.com/watch?v=c5aN5zOEpM8	1:09	12.04
clip8	https://www.youtube.com/watch?v=c5aN5zOEpM8	1:21	7.64
clip9	https://youtu.be/_npyQonU-V0?t=5m03s **(N/A)**	5:03	7.92
clip10	https://youtu.be/_npyQonU-V0?t=9m6s **(N/A)**	9:06	7.16
clip11	https://www.youtube.com/watch?v=c5aN5zOEpM8	1:01	6.36
clip12	https://www.youtube.com/watch?v=i6i1PpEtM_4	2:43	10.40
clip13	https://youtu.be/wT9Prne9wF0 **(N/A)**	0:00	6.40
clip14	https://youtu.be/_npyQonU-V0?t=31s **(N/A)**	0:31	8.04
clip15	https://www.youtube.com/watch?v=i6i1PpEtM_4	1:42	8.04
clip16	https://www.youtube.com/watch?v=i6i1PpEtM_4	2:35	3.68
clip17	https://www.youtube.com/watch?v=i6i1PpEtM_4	2:24	7.72
clip18	https://www.youtube.com/watch?v=i6i1PpEtM_4	2:52	4.40
clip19	https://youtu.be/iPDM0msZwQk?t=8s	0:08	10.24
clip20	https://youtu.be/_npyQonU-V0?t=1m8s **(N/A)**	1:08	4.80
clip21	https://youtu.be/_npyQonU-V0?t=3m50s **(N/A)**	03:50	5.24
clip22	https://youtu.be/T84se4fc4KU?t=24s	0:23	3.68
clip23	https://www.youtube.com/watch?v=i6i1PpEtM_4	0:45	2.76
clip24	https://www.youtube.com/watch?v=c5aN5zOEpM8	0:51	3.72
clip25	https://www.youtube.com/watch?v=JPhIPT9yOu8	0:42	8.00
clip26	https://youtu.be/WmuybcgSjkI **(N/A)**	3:24	3.20
clip27	https://youtu.be/_npyQonU-V0?t=47s **(N/A)**	0:47	3.60
clip28	https://youtu.be/09ZtoTthtys?t=4m50s	4:50	5.56
clip29	https://youtu.be/HVK-xbdddhA?t=3m8s	3:07	6.32
clip30	https://youtu.be/09ZtoTthtys?t=9m13s	9:14	3.20
clip31	https://youtu.be/V-mR66UW8NI?t=105	1:45	2.08
clip32	https://youtu.be/09ZtoTthtys?t=4m57s	4:57	3.00
clip33	https://www.youtube.com/watch?v=muN0yh_STkM	0:05	2.64
clip34	https://youtu.be/V-mR66UW8NI?t=30s	0:30	4.04

⁺As of publication, please note that some links no longer exist since they were collected in 2017. Due to copyright issues, we cannot upload the full videos publicly. We have indicated with N/A next to the links that do not exists. Please do not hesitate to contact the authors if interested in viewing the original video.

List of Motor and Social Words Classification in Experiment 1

Word	Category	Frequency
try	social	111
want	social	84
long	social	55
attempt	social	39
know	social	25
uncomfortable	social	23
confused	social	20
kiss	social	20
think	social	18
give	social	16
initiating	social	10
expecting	social	9
pull	motor	45
reach	motor	39
hold	motor	38
gesture	motor	38
grab	motor	36
change	motor	35
move	motor	23
touch	motor	20
pull	motor	20
times	motor	19
far	motor	19
towards	motor	19
turned	motor	17
timing	motor	11
respond	motor	11
playing	motor	11
extended	motor	10
continued	motor	9
switched	motor	9