Cognitive Characteristics and Neural Mechanisms of Social Interaction Processing: A Third-Person Perspective

Authors: Cheng Yuhui, Yuan Xiangyong, Jiang Yi, Yuan Xiangyong

Date: 2020-11-02T00:00:00+00:00

Abstract

In daily life, identifying and comprehending others' social interactions from a "third-person" perspective is of paramount importance. This social interaction processing exhibits two cognitive characteristics: configural holism and action association; it is manifested as a hierarchical processing procedure involving the participation of numerous brain regions, primarily encompassing the person perception network, action observation network, and mentalizing network. Among these, brain regions such as the posterior superior temporal sulcus play a pivotal role in representing social interaction relationships. Future research should integrate multiple technical methodologies to further elucidate the genetic properties and neural mechanisms underlying social interaction processing, while also attending to its practical applications in real-world contexts.

Full Text

The Cognitive Characteristics and Brain Mechanisms of Social Interaction Processing from a Third-Person Perspective

CHENG Yuhui, YUAN Xiangyong, JIANG Yi

(State Key Laboratory of Brain and Cognitive Science, CAS Center for Excellence in Brain Science and Intelligence Technology, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China)
(Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049, China)
(Chinese Institute for Brain Research, Beijing 102206, China)

Abstract: The ability to recognize and understand others' social interactions from a third-person perspective is essential in daily life. Social interaction processing is characterized by two cognitive features: configural integrity and action contingency, and involves a hierarchical process engaging multiple brain regions, primarily including the person perception network, action observation network, and mentalizing network. Among these, the posterior superior temporal sulus plays a key role in representing social interaction relationships. Future research should combine multiple techniques to further reveal the genetic characteristics and neural mechanisms of social interaction processing and explore its practical applications.

Keywords: social interaction; third-person perspective; configural processing; action contingency; brain mechanism

1 Introduction

Social interaction refers to a meaningful social exchange process in which individuals express communicative intentions through actions or language, and their partners understand these intentions and respond appropriately (Capozzi & Ristic, 2018; Hauser & Wood, 2010). Social interactions are ubiquitous in daily life, encompassing various forms such as chatting with friends, dancing with partners, and sparring with opponents. Research on social interaction can adopt two distinct perspectives: the second-person perspective and the third-person perspective. Second-person perspective studies typically create real-time interactive situations to examine cognitive processes when participants directly engage with others (Caruana, McArthur, Woolgar, & Brock, 2017; Schilbach et al., 2013). In contrast, third-person perspective research focuses on participants as observers, investigating cognitive processes when watching others interact (Quadflieg & Koldewyn, 2017), including perception of interactive partners' features and inference of their intentions. Observing and understanding others' social interactions is not only crucial for human survival and evolution but also plays a vital role in our daily lives. It provides opportunities for observational learning, helps individuals better understand others' personality traits and social relationships, deepens comprehension of social groups, and enhances social skills (Matheson, Moore, & Akhtar, 2013; Milinski, 2016).

In recent years, an increasing number of studies have explored social interaction from a third-person perspective, which can reveal the cognitive and neural mechanisms underlying human processing of others' interactions while providing references for diagnosing special populations and developing AI algorithms for social interaction recognition. Consequently, this line of research holds significant theoretical and applied value. This article adopts a third-person perspective to discuss the prerequisites and cognitive characteristics of social interaction processing, examines its cognitive and neural mechanisms, and identifies important questions for future investigation. Therefore, all subsequent mentions of social interaction specifically refer to third-person perspective social interaction.

2 Prerequisites for Social Interaction Processing

Social interactions can be presented in diverse formats, including brief videos (Sinke, Sorger, Goebel, & de Gelder, 2010), static images (Quadflieg, Gentile, & Rossion, 2015), point-light animations that abstract humans into joint points (Manera, Schouten, Becchio, Bara, & Verfaillie, 2010), and simple geometric figures exhibiting only biological motion patterns (Gao, Scholl, & McCarthy, 2012). By merely observing these stimuli, people can judge whether two parties are interacting, identify the specific type of interaction (Manera, von der Luhe, Schilbach, Verfaillie, & Becchio, 2016), and infer various social information such as whether they are friendly or hostile (Wu, Hua, Yang, & Yin, 2018), how intimate their relationship is (Costanzo & Archer, 1989), or whether they share the same social status (Floyd & Erbert, 2003; Mast & Hall, 2004).

Regardless of presentation format, two individuals are perceived as engaging in social interaction only when three prerequisite conditions are simultaneously satisfied across spatial, temporal, and meaningful dimensions. Spatially, both interpersonal distance and body orientation influence observers' perception of social interaction. Recent research has found that the closer two individuals are and the more they face each other, the more likely they are to be judged as interacting (Hartmann et al., 2019; Zhou, Han, Liang, Hu, & Kuai, 2019). Temporally, when two individuals' actions are continuous—that is, when Person A's action is promptly followed by Person B's response—they are more likely to be perceived as interacting. As the temporal interval between their actions increases, Person A's action becomes less predictive of Person B's action (Manera, Schouten, Verfaillie, & Becchio, 2013), indicating that observers' perception of the interaction diminishes. In terms of meaning, two individuals' actions must form a logically coherent connection to be perceived as socially interactive. For example, when Person A issues a "stand up" command and Person B rises from a squatting position, the two are perceived as interacting; however, if Person A issues the same command while Person B continues sweeping throughout, they are considered independent individuals without interaction (Manera et al., 2015). In summary, close proximity, facing orientation, temporal continuity, and logical meaning are all prerequisites for perceiving social interaction. Most researchers generate non-social interaction control stimuli based on one of these prerequisites, such as back-to-back body postures, scrambled temporal structures of interactive actions, or paired actions without logical association. However, in real social situations, these factors often jointly influence judgments of social interaction, yet only a handful of studies have simultaneously examined the role of multiple factors in social interaction perception (Hartmann et al., 2019). Future research could investigate how these factors work together from an integrated perspective.

3 Cognitive Characteristics of Social Interaction Processing

Once two interacting individuals are perceived by a third party as engaging in social interaction, they are processed as a holistic configural unit in space, while their actions form an interrelated and meaningful temporal structure. This configural integrity and action contingency constitute the cognitive characteristics of social interaction processing and represent potential sources of its processing advantages (Fedorov, Chang, Giese, Bulthoff, & de la Rosa, 2018; Neri, Luu, & Levi, 2006; Papeo, Goupil, & Soto-Faraco, 2019; Vestner, Tipper, Hartley, Over, & Rueschemeyer, 2019). The following sections elaborate on evidence supporting these two cognitive characteristics and their specific manifestations in processing advantages.

3.1 Holistic Configural Processing

Holistic configural processing of social interaction refers to the holistic representation of the spatial structure formed by two interacting individuals (face-to-face body orientation), wherein the interactive dyad is processed and stored as a structured functional unit rather than two independent individuals (Ding, Gao, & Shen, 2017; Papeo et al., 2019; Papeo, Stein, & Soto-Faraco, 2017; Vestner et al., 2019). If social interaction indeed involves holistic configural processing, then the interactive dyad as a unit should influence spatial attention allocation (Shen, Yin, Ding, Shui, & Zhou, 2016; Yin, Xu, Duan, & Shen, 2018), exhibiting the classic object-based attentional effect—where attentional resources are automatically allocated to different parts of the same object, yielding faster responses to them than to parts of other objects. In social interaction contexts, when a cue appears on one person, participants respond faster not only to targets on that same person but also to targets on the interacting partner compared to targets on non-interacting individuals; this effect disappears in non-interactive scenarios (Ji, Yin, Huang, & Ding, in press). Moreover, holistic configural processing is directly reflected in the phenomenon that the subjective distance between interacting individuals is perceived as closer than their actual physical distance (Vestner et al., 2019).

In holistic configural processing, the spatial structure of interactive partners is represented as a unit. Compared to piecemeal processing of local spatial information, this holistic approach reduces cognitive load to some extent, thereby conferring processing advantages to interactive over non-interactive stimuli (Papeo et al., 2019; Papeo et al., 2017; Vestner, Gray, & Cook, 2020). First, holistic configural processing facilitates rapid access of interactive stimuli to conscious awareness; participants' accuracy in identifying masked, subthreshold interactive figures facing each other is significantly higher than for masked, non-interactive back-to-back figures, while this recognition advantage disappears for inverted stimuli (Papeo et al., 2017). Second, holistic configural processing also expedites visual search for interactive stimuli, demonstrating search asymmetry: searching for face-to-face figures among back-to-back figures yields higher accuracy and faster responses than searching for back-to-back figures among face-to-face figures (Papeo et al., 2019). Finally, holistic configural processing benefits memory storage and retrieval. Participants show higher accuracy for interactive than non-interactive stimuli in both short-term and long-term memory tasks (Ding et al., 2017; Vestner et al., 2019). These studies not only demonstrate the processing advantages conferred by holistic configural processing but also corroborate the existence of holistic configural representations in social interaction processing. However, whether holistic configural processing is specific to social interaction stimuli or also occurs in other face-to-face objects remains unresolved (Papeo & Abassi, 2019; Vestner et al., 2020). If holistic configural processing is specific to social interaction stimuli, whether it applies to all types of social interactions (e.g., cooperative or competitive) also awaits further investigation (Yin et al., 2013).

3.2 Contingent Action Representation

Social interaction is not only processed as a holistic unit in space but also constitutes a dynamic, continuous temporal process. In social interaction, whether it involves a social initiator conveying communicative intentions and a responder reacting (e.g., Person A throws an object, Person B catches it) or interactive partners performing coordinated joint actions through 默契配合 (such as partner dancing; Marsh, Richardson, & Schmidt, 2009; Pesquita, Whitwell, & Enns, 2018), the partners' actions are interconnected and form a meaningful, continuous temporal structure, resulting in interactive actions being represented in the brain as contingently related. Researchers have used two pairs of interactive actions—where "receiving" responds to "giving" and "catching" responds to "throwing" —to test whether observing one action automatically triggers representations of its associated action. They found that repeatedly presenting the "receiving" action made participants more likely to perceive an ambiguous action (resembling both "receiving" and "catching") as "catching" ; crucially, even after repeated presentation of the "giving" action, participants still perceived the ambiguous action as "catching." Thus, despite substantial visual differences between "giving" and "receiving" actions, they produced identical adaptation effects. This cross-adaptation effect provides compelling evidence for contingent action representations (Fedorov et al., 2018). Additional evidence for contingent action representation includes findings that participants perceive shorter temporal durations (Liu, Yuan, Chen, Jiang, & Zhou, 2018) and higher action fluency (Peng, Ichien, & Lu, 2020) for interactive actions.

Contingent action representation enables human perception of others' interactive behavior to rely not only on current visual input but also on prior knowledge and the continuous temporal structure between actions to predict upcoming actions, thereby facilitating interactive action recognition under high uncertainty (Manera, Becchio, Schouten, Bara, & Verfaillie, 2011; Manera, Giudice, Bara, Verfaillie, & Becchio, 2011; Neri et al., 2006; von der Luhe et al., 2016). Neri and colleagues (2006) first demonstrated this facilitatory effect by presenting participants with two sequentially noise-masked point-light animations, one containing two interacting or non-interacting figures and the other containing only one figure. They found that participants achieved higher accuracy in judging the two-figure animation when it was a temporally complete interactive animation. Subsequent research confirmed that participants could use the initiator' s action to predict the responder' s action: when the responder (Person B) in a two-person animation was masked by noise, participants more easily detected Person B from the noise if Persons A and B had an interactive relationship (Manera, Becchio, et al., 2011; Manera, Giudice, et al., 2011; Manera et al., 2013). This predictive effect based on contingent actions has also been validated using binocular rivalry paradigms (Su, Van Boxtel, & Lu, 2016). Despite employing different experimental paradigms and diverse stimulus formats, these studies converge on the conclusion that contingent action representation facilitates interactive action recognition. However, current research has primarily focused on whole-body actions, with less attention to local limb movements such as gestures and foot actions (Zaini, Fawcett, White, & Newman, 2013). In certain contexts like dancing, local limb movements serve as important information carriers; future research could employ more ecologically valid stimuli to explore the role of contingent representations of local limb movements in social interaction.

4 Brain Mechanisms of Social Interaction Processing

As a high-level cognitive activity, social interaction processing constitutes a hierarchical process that progresses from recognizing bodies and faces to understanding interactive actions and inferring communicative intentions. This process can be divided into three stages corresponding to three brain networks (see Figure 1 [Figure 1: see original paper]): the person perception network, action observation network, and mentalizing network (see Quadflieg & Koldewyn, 2017 for review). The person perception network performs visual analysis of interactors' bodies and faces, primarily including the fusiform face area (FFA) specialized for face processing, the extrastriate body area (EBA) for body processing, and the posterior superior temporal sulcus (pSTS) for dynamic faces and biological motion (Quadflieg et al., 2015). The action observation network, part of the mirror neuron system (MNS), participates in rapid action recognition and accurate intention understanding, comprising the premotor cortex (PMC), inferior frontal gyri (IFG), and inferior parietal lobule (IPL) (Caspers, Zilles, Laird, & Eickhoff, 2010). The mentalizing network forms the neural basis of theory of mind, primarily responsible for inferring others' abstract mental states (e.g., desires, motivations, intentions, beliefs) and complex social attributes (e.g., social norms, kinship, emotional states; Keysers & Gazzola, 2007), including brain regions such as dorsal/ventral medial prefrontal cortex (dmPFC/vmPFC), temporoparietal junction (TPJ), and precuneus (PrC) (Schurz, Radua, Aichhorn, Richlan, & Perner, 2014).

Figure 1. Brain networks in social interaction processing. Each brain region corresponds to the brain network of the same color below.

4.1 Brain Regions Related to Cognitive Characteristics of Social Interaction

When investigating the brain networks underlying social interaction processing, researchers initially focused on brain regions activated when viewing interactive stimuli. However, with developments in behavioral paradigms and data analysis methods, more researchers have begun identifying brain regions specifically involved in processing the cognitive characteristics of social interaction. First, are there specific brain regions that support holistic configural processing of social interaction? Existing evidence suggests that the EBA, part of the person perception network, is sensitive to body posture information and encodes actions based on this information (Downing, Peelen, Wiggett, & Tew, 2006), making it a candidate region for holistic configural processing of social interaction. To test this hypothesis, researchers used multivoxel pattern analysis (MVPA) to train classifiers to discriminate types of interactive videos (arguing, celebrating, or laughing) based on activation patterns in pSTS or EBA, then tested performance on discriminating two-person interactive actions versus single-person actions. Results showed that while pSTS could distinguish between interactive and single-person actions, accuracy did not differ significantly between the two conditions. In contrast, EBA showed significantly higher accuracy for discriminating interactive actions than single-person actions, indicating that EBA extracts unique holistic information about social interaction beyond processing local single-person actions (Walbrin & Koldewyn, 2019). Subsequent research further demonstrated that EBA' s holistic encoding is limited to face-to-face figures: after training a classifier to discriminate actions of a single figure, identification accuracy for that figure' s action was higher when the figure faced another figure versus when they stood back-to-back. That is, when two figures faced each other, forming a social interaction relationship, EBA' s ability to encode a single figure' s body was enhanced (Abassi & Papeo, 2020). These results demonstrate that EBA plays an important role in holistic configural processing of social interaction (Abassi & Papeo, 2020; Walbrin & Koldewyn, 2019).

Second, researchers have also examined whether certain brain regions represent contingent actions in social interaction. Since contingent action representation involves temporal continuity and meaningful association of interactive actions, it may involve both the action observation and mentalizing networks. In social interaction, contingent actions (compared to non-contingent actions where one interactors' actions are replaced by mirror images) more strongly activate the action observation network, including IFG (extending to PMC), pSTS, and left IPL (Georgescu et al., 2014). Additionally, Zillekens et al. (2019) used the noise-masking paradigm developed previously (Manera, Giudice, et al., 2011) to preliminarily investigate the neural mechanisms of predicting others' actions in social interaction. By manipulating whether Person A conveyed communicative intentions to Person B, they created interactive versus non-interactive animations, masked Person B with noise, and asked participants to judge whether Person B was present. Results showed that regardless of Person B' s presence, interactive animations significantly enhanced activation in superior frontal gyrus (SFG) and activated superior parietal lobule (SPL) and inferior temporal gyri (ITG) within the action observation network more than non-interactive animations. These findings suggest that in interactive contexts, the brain can form expectations about Person B' s actions based on Person A' s actions, reducing demands on executive control and consequently showing decreased activation in the action observation network. When expectations were violated—when Person B was absent in interactive animations or present in non-interactive animations—right posterior medial frontal gyrus (rPMFG), bilateral precuneus, left cerebellum, and left fusiform gyrus showed stronger activation, possibly reflecting prediction errors in social interaction. Furthermore, amygdala activation predicted participants' ability to discriminate Person B, and functional connectivity between amygdala and mPFC was enhanced under interactive conditions, reflecting that expectations about interactive actions involve complex social inference that may require integration of the mentalizing network (Zillekens et al., 2019).

These studies have identified brain regions corresponding to the cognitive characteristics of social interaction, providing initial directions for uncovering the neural mechanisms of holistic configural processing and contingent action representation. However, many important questions remain unanswered, such as whether the activation pattern of an interaction initiator' s action can be used to decode the responder' s action. Moreover, research urgently needs to establish links between behavioral task performance and neuroimaging results to more accurately reveal the relationship between behavior and brain function—for instance, whether EBA activation can predict performance on holistic configural processing tasks and reflect individual differences in perceiving social interaction.

4.2 Specific Brain Regions for Social Interaction Processing

Although social interaction processing activates numerous brain regions across different stages, some regions may only participate in single-person body perception and action understanding. Therefore, researchers have attempted to isolate brain regions that specifically represent interaction relationships through more refined experimental designs. In a macaque study using functional magnetic resonance imaging (fMRI), researchers examined whole-brain activation while monkeys watched different interactive videos (Sliwa & Freiwald, 2017). Results revealed that only medial prefrontal regions and parts of IPL (corresponding to the human mentalizing network) were specifically sensitive to monkey-monkey social interaction, showing no response to videos of monkey-object or object-object interactions. This suggests that specific brain regions representing social interaction relationships exist in the macaque brain. But do similar regions exist in the human brain? Researchers have focused on pSTS, which is considered part of not only the person perception network but also the action observation and mentalizing networks (Deen, Koldewyn, Kanwisher, & Saxe, 2015; Yang, Rosenblau, Keifer, & Pelphrey, 2015), and responds to various types of social stimuli, earning it the designation as a social brain "hub" (Lahnakoski et al., 2012). To further exclude confounding variables, Isik et al. (2017) simplified the physical properties of social interaction stimuli, retaining only their interaction relationships. They found that interactive animations composed of simple geometric figures (two figures helping or hindering each other) more strongly activated pSTS than physical interactions (two figures colliding) or single-person animations. Moreover, using MVPA, they could discriminate whether the interaction relationship was helping or hindering based on pSTS activation patterns. This experiment demonstrated that pSTS can represent social interaction relationships (see similar results in Walbrin, Downing, & Koldewyn, 2018) and may be a specific brain region for processing social interaction (Isik, Koldewyn, Beeler, & Kanwisher, 2017). Notably, however, pSTS was only the most strongly responding region to social interaction in this study, with its boundaries extending to nearby theory-of-mind regions (TPJ) and functionally localized dynamic face regions. In summary, specific brain regions that truly represent social interaction relationships may exist in the brain—namely, the mentalizing network and adjacent pSTS (Isik et al., 2017)—corresponding to the processing stage of understanding partners' intentions and mental states.

4.3 Functional Connectivity in Social Interaction Processing

In recent years, some researchers have examined information transfer and functional connectivity among different brain regions from a hierarchical processing perspective. A meta-analysis of neuroimaging studies indicated that during social interaction processing, the pathway from recognizing individual actions to parsing intentions of single and interactive actions to inferring mental states behind action intentions primarily activates a route from middle-posterior temporal cortex through pSTS to TPJ (Arioli & Canessa, 2019). Among these regions, pSTS likely serves as a crucial node in social interaction processing, performing information classification and transmission to other brain regions (Sliwa & Freiwald, 2017). Therefore, functional connectivity between pSTS and other brain regions may also be modulated by social interaction type. When understanding shared action intentions, such as two people cooperating to move a box, pSTS enhances connectivity with superior parietal lobule (SPL) and ventral PMC in the action observation network, whereas when understanding emotional interactions, such as two people holding hands and smiling, pSTS enhances connectivity with vmPFC in the mentalizing network (Arioli et al., 2018). Currently, research on functional connectivity is only in its infancy, and future studies require more data to comprehensively characterize the functional connectivity patterns underlying different stages and types of social interaction processing and to reveal the role of key nodes like pSTS in social interaction information transmission.

5 Summary and Outlook

In social environments, others' social interactions provide rich social information, making detection and recognition of social interaction cues critically important. This importance renders humans highly sensitive to social interaction, similar to faces and biological motion, and confers processing advantages: social interaction stimuli are prioritized over non-social interaction stimuli, and the ability to identify and remember social interaction stimuli is enhanced. The deep sources of these advantages lie in humans' holistic configural processing and contingent action representation of social interaction. Social interaction constitutes a hierarchical processing process involving brain networks including the person perception network, action observation network, and mentalizing network (Quadflieg & Koldewyn, 2017). Regarding the cognitive characteristics of social interaction processing, preliminary studies suggest that EBA participates in holistic configural processing, while some regions in the action observation and mentalizing networks are involved in contingent action representation. Although certain brain regions show greater sensitivity to social interaction relationships, information transfer among brain regions across different processing stages and corresponding functional connectivity are equally important for completing social interaction processing (Arioli & Canessa, 2019). In summary, despite extensive exploration of social interaction processing, several questions remain unresolved.

First, although humans can rapidly detect social interaction signals and correctly interpret others' action meanings, whether this ability is innately determined remains uncertain. Research shows that infants aged 5–14 months prefer social interaction information, exhibiting longer looking times at interactive than non-interactive animations when both are presented simultaneously (Galazka, Roche, Nystrom, & Falck-Ytter, 2014), and in dynamic chasing animations, they look significantly longer at the "chaser" than at the "fleeing" or non-interactive figures (Galazka & Nystrom, 2016). These findings indicate that infants under one year already show greater interest in interactive information and interaction-initiating individuals in their environment, suggesting this ability may have been preserved through natural selection during human evolution and thus possesses some heritability. Future research could employ behavioral genetics methods, such as twin studies (Wang et al., 2018), to provide more direct evidence for the heritability of social interaction processing.

Second, numerous studies have demonstrated that humans exhibit holistic processing advantages for stimuli conveying substantial biological and social information, such as faces and biological motion. This processing depends on spatial structural relationships among parts and shows classic inversion effects (Farah, Tanaka, & Drain, 1995; Reed, Stone, Bozova, & Tanaka, 2003). Compared to fine contour lines and other high spatial frequency information, coarse holistic structure is reflected in low spatial frequency information and is processed faster than high-frequency information. Consequently, this holistic processing advantage is generally considered to exist only for low-frequency stimulus information (Goffaux et al., 2011). Similar to faces and biological motion, social interaction is processed configurally and conveys important social information with evolutionary significance. Therefore, the holistic configural processing of social interaction may be specific to low-frequency components of interactive stimuli, thereby conferring processing advantages over non-interactive stimuli. However, this inference requires further experimental investigation.

Third, research on neural mechanisms has established social interaction processing as a hierarchical process. Although existing studies have revealed the roles of the action observation network, mentalizing network, and pSTS in social interaction processing (Arioli & Canessa, 2019; Sliwa & Freiwald, 2017), many questions remain from the perspectives of neural mechanisms of cognitive characteristics, specific brain regions, and functional connectivity among brain regions. For example, does functional connectivity among relevant brain regions change dynamically with the progression of interaction when predicting others' actions? Does pSTS function to filter and transmit social interaction information? How similar or different are the brain mechanisms across different types of social interaction (Canessa et al., 2012; Sinke et al., 2010), and do they depend on different functional connectivity patterns (Arioli et al., 2018)? Finally, although this article focuses on third-person social interaction processing, with the development of hyperscanning technology for simultaneously recording neural activity from multiple brains, the neural mechanisms of real-time human interaction are becoming a central focus of second-person social interaction research (Kingsbury & Hong, 2020). However, few studies have directly compared social interaction processing between these two perspectives, and future research could combine multiple techniques to explore similarities and differences in their underlying brain mechanisms.

Finally, current research has primarily focused on the cognitive and neural mechanisms of social interaction processing, but future work should explore its applied value more extensively. Multiple lines of evidence indicate that social interaction processing abilities are impaired in certain clinical populations (Centelles, Assaiante, Etchegoyhen, Bouvard, & Schmitz, 2013; Kuschefski, Falter-Wagner, Bente, Vogeley, & Georgescu, 2019; Okruszek et al., 2015; Walter et al., 2009). For example, individuals with high-functioning autism cannot use action contingency to predict others' actions (von der Luhe et al., 2016), while patients with schizophrenia, although capable of representing contingent actions, show reduced ability to recognize biological motion in social interactions (Okruszek, Piejka, Wysokinski, Szczepocka, & Manera, 2018, 2019). Future research could employ more ecologically valid virtual reality technology to explore the cognitive characteristics of social interaction processing in these clinical populations and provide new insights for diagnosis and intervention. Additionally, using artificial intelligence to recognize human interactions in real-world scenarios, such as detecting suspicious groups in public spaces, holds important and broad application prospects.

References

Abassi, E., & Papeo, L. (2020). The representation of two-body shapes in the human visual cortex. *The Journal of Neuroscience, 40*(4), 852-863.

Arioli, M., & Canessa, N. (2019). Neural processing of social interaction: Coordinate-based meta-analytic evidence from human neuroimaging studies. *Human Brain Mapping, 40*(13), 3712-3737.

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.

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

Capozzi, F., & Ristic, J. (2018). How attention gates social interactions. *Annals of the New York Academy of Sciences, 1426*(1), 179-198.

Caruana, N., McArthur, G., Woolgar, A., & Brock, J. (2017). Simulating social interactions for the experimental investigation of joint attention. *Neuroscience & Biobehavioral Reviews, 74*, 115-125.

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

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

Costanzo, M., & Archer, D. (1989). Interperting the expressive behavior of others: The interpersonal perception task. *Journal of Nonverbal Behavior, 13*(4), 225-245.

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

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

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.

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance, 21*(3), 628-634.

Fedorov, L. A., Chang, D. S., Giese, M. A., Bulthoff, H. H., & de la Rosa, S. (2018). Adaptation aftereffects reveal representations for encoding of contingent social actions. *Proceedings of the National Academy of Sciences, 115*(29), 7515–7520.

Floyd, K., & Erbert, L. A. (2003). Relational message interpretations of nonverbal matching behavior: An application of the social meaning model. *The Journal of Social Psychology, 143*(5), 581–597.

Galazka, M., & Nystrom, P. (2016). Infants' preference for individual agents within chasing interactions. *Journal of Experimental Child Psychology, 147*, 53–70.

Galazka, M. A., Roche, L., Nystrom, P., & Falck-Ytter, T. (2014). Human infants detect other people's interactions based on complex patterns of kinematic information. *PLoS ONE, 9*(11), e112432.

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

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

Goffaux, V., Peters, J. C., Haubrechts, J., Schiltz, C., Jansma, B., & Goebel, R. (2011). From coarse to fine? Spatial and temporal dynamics of cortical face processing. *Cerebral Cortex, 21*(2), 467–476.

Hartmann, M., Mavrolampados, A., Allingham, E., Carlson, E., Burger, B., & Toiviainen, P. (2019). Kinematics of perceived dyadic coordination in dance. *Scientific Reports, 9*(1), 15594.

Hauser, M., & Wood, J. (2010). Evolving the capacity to understand actions, intentions, and goals. *Annual Review of Psychology, 61*(1), 303–324.

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.

Ji, H., Yin, J., Huang, Y., & Ding, X. (in press). Selective attention operates on the group level for interactive biological motion. *Journal of Experimental Psychology: Human Perception and Performance*.

Keysers, C., & Gazzola, V. (2007). Integrating simulation and theory of mind: from self to social cognition. *Trends in Cognitive Sciences, 11*(5), 194–196.

Kingsbury, L., & Hong, W. (2020). A multi-brain framework for social interaction. *Trends in neurosciences, 43*(9), Kuschefski, M., Falter-Wagner, C. M., Bente, G., Vogeley, K., & Georgescu, A. L. (2019). Inferring power and dominance from dyadic nonverbal interactions in autism spectrum disorder. *Autism Research, 12*(3), 505–

Lahnakoski, J. M., Glerean, E., Salmi, J., Jaaskelainen, I. P., Sams, M., Hari, R., & Nummenmaa, L. (2012). Naturalistic fMRI mapping reveals superior temporal sulcus as the hub for the distributed brain network for social perception. *Frontiers in Human Neuroscience, 6*(8), 233.

Liu, R., Yuan, X., Chen, K., Jiang, Y., & Zhou, W. (2018). Perception of social interaction compresses subjective duration in an oxytocin-dependent manner. *Elife, 7*, e32100.

Manera, V., Becchio, C., Schouten, B., Bara, B. G., & Verfaillie, K. (2011). Communicative interactions improve visual detection of biological motion. *PLoS ONE, 6*(1), e14594.

Manera, V., Giudice, M., Bara, B. G., Verfaillie, K., & Becchio, C. (2011). The second-agent effect: Communicative gestures increase the likelihood of perceiving a second agent. *PLoS ONE, 6*(7), e22650.

Manera, V., Iani, F., Bourgeois, J., Haman, M., Okruszek, L. P., Rivera, S. M., . . . Becchio, C. (2015). The Multilingual CID-5: A new tool to study the perception of communicative interactions in different languages. *Frontiers in Psychology, 6*, 1724.

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

Manera, V., Schouten, B., Verfaillie, K., & Becchio, C. (2013). Time will show: Real time predictions during interpersonal action perception. *PLoS ONE, 8*(1), e54949.

Manera, V., von der Luhe, T., Schilbach, L., Verfaillie, K., & Becchio, C. (2016). Communicative interactions in point-light displays: Choosing among multiple response alternatives. *Behavior Research Methods, 48*(4),

Marsh, K. L., Richardson, M. J., & Schmidt, R. C. (2009). Social connection through joint action and interpersonal coordination. *Topics in Cognitive Science, 1*(2), 320–339.

Mast, M. S., & Hall, J. A. (2004). Who is the boss and who is not? Accuracy of judging status. *Journal of Nonverbal Behavior, 28*(3), 145–165.

Matheson, H., Moore, C., & Akhtar, N. (2013). The development of social learning in interactive and observational contexts. *Journal of Experimental Child Psychology, 114*(2), 161–172.

Milinski, M. (2016). Reputation, a universal currency for human social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*(1687), 20150100.

Neri, P., Luu, J. Y., & Levi, D. M. (2006). Meaningful interactions can enhance visual discrimination of human agents. *Nature Neuroscience, 9*(9), 1186–1192.

Okruszek, L., Haman, M., Kalinowski, K., Talarowska, M., Becchio, C., & Manera, V. (2015). Impaired recognition of communicative interactions from biological motion in schizophrenia. *PLoS ONE, 10*(2), e0116793.

Okruszek, L., Piejka, A., Wysokinski, A., Szczepocka, E., & Manera, V. (2018). Biological motion sensitivity, but not interpersonal predictive coding is impaired in schizophrenia. *Journal of Abnormal Psychology, 127*(3),

Okruszek, L., Piejka, A., Wysokinski, A., Szczepocka, E., & Manera, V. (2019). The second agent effect: Interpersonal predictive coding in people with schizophrenia. *Social Neuroscience, 14*(2), 208–213.

Papeo, L., & Abassi, E. (2019). Seeing social events: The visual specialization for dyadic human-human interactions. *Journal of Experimental Psychology: Human Perception and Performance, 45*(7), 877–888.

Papeo, L., Goupil, N., & Soto-Faraco, S. (2019). Visual search for people among people. *Psychological Science, 30*(10), 1483–1496.

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

Peng, Y., Ichien, N., & Lu, H. (2020). Causal actions enhance perception of continuous body movements. *Cognition, 194*, 104060.

Pesquita, A., Whitwell, R. L., & Enns, J. T. (2018). Predictive joint-action model: A hierarchical predictive approach to human cooperation. *Psychonomic Bulletin & Review, 25*(5), 1751–1769.

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

Quadflieg, S., & Koldewyn, K. (2017). The neuroscience of people watching: How the human brain makes sense of other people' s encounters. *Annals of the New York Academy of Sciences, 1396*(1), 166–182.

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

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

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

Shen, M., Yin, J., Ding, X., Shui, R., & Zhou, J. (2016). Deployment of attention on handshakes. *Frontiers in Psychology, 7*, 681.

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

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

Su, J., Van Boxtel, J. J., & Lu, H. (2016). Social interactions receive priority to conscious perception. *PLoS ONE, 11*(8), e0160468.

Vestner, T., Gray, K. L. H., & Cook, R. (2020). Why are social interactions found quickly in visual search tasks? *Cognition, 200*, 104270.

Vestner, T., Tipper, S. P., Hartley, T., Over, H., & Rueschemeyer, S.A. (2019). Bound together: Social binding leads to faster processing, spatial distortion, and enhanced memory of interacting partners. *Journal of Experimental Psychology: General, 148*(7), 1251-1268.

von der Luhe, T., Manera, V., Barisic, I., Becchio, C., Vogeley, K., & Schilbach, L. (2016). Interpersonal predictive coding, not action perception, is impaired in autism. *Philosophical Transactions of the Royal Society B Biological Sciences, 371*(1693).

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

Walbrin, J., & Koldewyn, K. (2019). Dyadic interaction processing in the posterior temporal cortex. *Neuroimage, 198*, 296-302.

Walter, H., Ciaramidaro, A., Adenzato, M., Vasic, N., Ardito, R. B., Erk, S., & Bara, B. G. (2009). Dysfunction of the social brain in schizophrenia is modulated by intention type: An fMRI study. *Social Cognitive and Affective Neuroscience, 4*(2), 166-176.

Wang, Y., Wang, L., Xu, Q., Liu, D., Chen, L., Troje, N. F., . . . Jiang, Y. (2018). Heritable aspects of biological motion perception and its covariation with autistic traits. *Proceedings of the National Academy of Sciences, 115*(8), 1937-1942.

Wu, X., Hua, R., Yang, Z., & Yin, J. (2018). The influence of intention and outcome on evaluations of social interaction. *Acta Psychologica, 182*, 75-81.

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

Yin, J., Ding, X., Zhou, J., Shui, R., Li, X., & Shen, M. (2013). Social grouping: Perceptual grouping of objects by cooperative but not competitive relationships in dynamic chase. *Cognition, 129*(1), 194-204.

Yin, J., Xu, H., Duan, J., & Shen, M. (2018). Object-based attention on social units: Visual selection of hands performing a social interaction. *Psychological*

Science, 29(7), 1040-1048.

Zaini, H., Fawcett, J. M., White, N. C., & Newman, A. J. (2013). Communicative and noncommunicative point-light actions featuring high-resolution representation of the hands and fingers. *Behavior Research Methods, 45*(2), 319-328.

Zhou, C., Han, M., Liang, Q., Hu, Y. F., & Kuai, S. G. (2019). A social interaction field model accurately identifies static and dynamic social groupings. *Nature Human Behaviour, 3*(8), 847-855.

Zillekens, I. C., Brandi, M.L., Lahnakoski, J. M., Koul, A., Manera, V., Becchio, C., & Schilbach, L. (2019). Increased functional coupling of the left amygdala and medial prefrontal cortex during the perception of communicative point-light stimuli. *Social Cognitive and Affective Neuroscience, 14*(1), 97-107.

Note: Figure translations are in progress. See original paper for figures.