SENSORIMOTOR FOUNDATIONS OF SOCIAL COGNITION

EDITED BY: Andreas K. Engel, Paul F. M. J. Verschure, Danica Kragic,
Daniel Polani, Alfred Oliver Effenberg and Peter König

PUBLISHED IN: Frontiers in Human Neuroscience, Frontiers in Psychology and
Frontiers in Robotics and AI

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88976-399-3
DOI 10.3389/978-2-88976-399-3

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

SENSORIMOTOR FOUNDATIONS OF SOCIAL COGNITION

Topic Editors:
Andreas K. Engel, University Medical Center Hamburg-Eppendorf, Germany
Paul F. M. J. Verschure, Radboud University Nijmegen, Netherlands
Danica Kragic, Royal Institute of Technology, Sweden
Daniel Polani, University of Hertfordshire, United Kingdom
Alfred Oliver Effenberg, Leibniz University Hannover, Germany
Peter König, Osnabrück University, Germany

Citation: Engel, A. K., Verschure, P. F. M. J., Kragic, D., Polani, D., Effenberg, A. O., König, P., eds. (2022). Sensorimotor Foundations of Social Cognition. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-399-3

Table of Contents

frontiers | Frontiers in Human Neuroscience

EDITORIAL
published: 08 July 2022
doi: 10.3389/fnhum.2022.971133

Editorial: Sensorimotor Foundations of Social Cognition

Andreas K. Engel[1]*, Paul F. M. J. Verschure[2], Danica Kragic[3], Daniel Polani[4], Alfred O. Effenberg[5] and Peter König[6]

[1] Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, [2] Synthetic Perceptive Emotive Cognitive Systems Laboratory, Fundació Institut de Bioenginyería de Catalunya, Barcelona, Spain, [3] School of Electrical Engineering and Computer Science, Royal Institute of Technology, Stockholm, Sweden, [4] Centre for Computer Science and Informatics Research, School of Computer Science, University of Hertfordshire, Hertfordshire, United Kingdom, [5] Institute of Sports Science, Leibniz Universität Hannover, Hannover, Germany, [6] Institute of Cognitive Science, Osnabrück University, Osnabrück, Germany

Keywords: social cognition, sensorimotor coupling, coordination, embodied cognition, human-robot interaction

OPEN ACCESS

Edited and reviewed by:
Lutz Jäncke,
University of Zurich, Switzerland

****Correspondence:***
Andreas K. Engel
ak.engel@uke.de

Specialty section:
This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

Received: *16 June 2022*
Accepted: *22 June 2022*
Published: *08 July 2022*

Citation:
Engel AK, Verschure PFMJ, Kragic D, Polani D, Effenberg AO and König P (2022) Editorial: Sensorimotor Foundations of Social Cognition. Front. Hum. Neurosci. 16:971133. doi: 10.3389/fnhum.2022.971133

Editorial on the Research Topic

Sensorimotor Foundations of Social Cognition

In classical representation-oriented approaches of social cognition, agents are thought to interact with conspecifics based on their capacity to develop a "theory-of-mind", i.e., to generate complex models of the intentions, beliefs, and personalities of their interaction partners. In this framework, the primary mode of interaction with the social environment is that of a detached observer who theorizes and produces inferences about other participants. In contrast, this Research Topic seeks to turn the spotlight on the grounding of social cognition in dynamic sensorimotor and informational coupling of agents, in human-human as well as human-robot interaction settings. According to this view, interaction dynamics hold substantial clues to the mechanism of social understanding and its disturbances (as for example observed in autism spectrum disorders). The argument is that high-level social deficits may be rooted in the impaired capacity for entraining and sustaining sensorimotor and informational coupling. Beyond novel insights into the mechanisms of functional and dysfunctional social behavior, the investigation of basic sensorimotor interaction patterns may help the development of socially competent robot technology. Tapping into the same logic, robotic agents sensitive to interpersonal sensorimotor contingencies should have an advantage over technology that does not consider this key aspect of human interaction. This Research Topic provides an interdisciplinary overview of trends and recent developments in conceptual, methodological and basic research, as well as applications of sensorimotor approaches in social cognitive science, neuroscience, and robotic research. One of the key questions is how concepts and methods from social cognitive and neuroscience transfer to human-robot interaction.

CONCEPTS

Several papers of this Research Topic discuss sensorimotor foundations of social cognition at the conceptual level, reviewing prevailing theories in the field and highlighting relevant experimental evidence that supports such theories. Lübbert et al. propose to extend the sensorimotor contingency theory into an action-oriented account of social cognition. The authors suggest that both informational and sensorimotor coupling between agents can support the use of action-effect contingencies in social context. The paper reviews the results of empirical studies that support the notion of socially shared sensorimotor contingencies, and discusses potential implications of this view for a better understanding of disturbed social interaction and for improvement of human-robot interactions. Vesper and Sevdalis also discuss possible functions of sensorimotor interactions in social context. They highlight three possible functions of sensorimotor communication, including the transfer of information on action intentions, the facilitation of predictions in joint action, and the stimulation of emotional experiences. A short opinion paper by Araneda Hinrichs criticizes the traditional view in social cognition research and emphasizes the usefulness of embodied and action-oriented concepts that seem more appropriate to account for social cognitive affordances. Along similar lines, an opinion piece by Rojas-Líbano and Parada argues for a key role of body-world coupling in the ontogeny of social cognition. The authors emphasize that both internal and external factors need to be considered in the modeling of social interactions by approaches from network science and by machine learning techniques. The paper by Tzafestas is a conceptual contribution reflecting upon general functions of imitation with a discussion of the nexus of the three concepts "imitation", "association", and "communicative function". To this end, the long-term formation of imitation by evolution is discussed in relation to the development of basic social and communicative skills.

SOCIAL COORDINATION DYNAMICS

This Research Topic comprises a number of behavioral studies on social coordination dynamics in a variety of different experimental paradigms. A study by Barone et al. used the perceptual crossing paradigm to develop a Turing test exploring the amount of minimally required information in terms of short-latency reciprocal sensorimotor contingencies—for human agents as well as for artificial agents. The study indicates that artificial agents should be able to generate short time reciprocal contingencies to make interactions with humans more fluid and, thus, more human-like. Using a similar paradigm, the study of Froese et al. investigated the short time emergence of sensorimotor contingencies based on haptic information by detecting the co-actor and perceptual learning of how changes in others' movements depend on changes in one's own movements, which is labeled as the "mastery of self-other contingencies". Two related studies by Jording and coworkers address the role of gaze cues in social interaction. The first of the studies (Jording, Engemann, et al.) investigated the gaze cues that lead an observer to ascribe a social intention to the observed person's gaze. The second study (Jording, Hartz et al.) used an interactive setup for the investigation of social gaze cues, where subjects had to judge whether the other person was trying to interact with them. The results from these studies support the pivotal role of gaze in social coordination and relationship formation and show that social expectations are reflected in differential responses to gaze patterns. Trendafilov et al. used a shared task that required coordinated action by two participants, who had to move an object by jointly tilting a tablet. The authors then used transfer entropy between the participants' actions to identify leader-follower relationships. The results show that transfer entropy is sufficiently sensitive to detect leader-follower directions and, furthermore, that leader-follower relations emerge spontaneously from the interaction, without being prompted. A study by Wahn et al. asked to which degree the task setting, i.e., individual vs. joint action, can influence sensory processing. Complementary to a previous report on the effect of joint performance of a crossmodal congruency task, the authors investigated a motion-discrimination and a temporal-order judgement task. All aspects relating to single subject performance are fully compatible with previous studies. However, they did not observe an influence of joint vs. single setups on performance in the two tasks. This demonstrates that at least a part of multisensory processing is performed independently of the social setting.

NEURAL AND AUTONOMIC SIGNATURES

Grounding of social cognition in the sensorimotor and informational coupling of agents has direct implications for any investigation of the physiological substrate. Foremost, it requires measurements of the physiological signals of interacting agents. This allows relating brain activity to the types of actions performed. Recording the physiological signals from both interacting agents simultaneously (hyperscanning) also allows relating both agents' neuronal activity to each other. Compared to classical single-subject laboratory setups, the step to recording multiple interacting participants entails substantial modifications. Czeszumski et al. present a review of methods, analysis techniques and results of recent hyperscanning research. Specifically, although most available physiological recording techniques have been used in hyperscanning setups, EEG and fNIRS see the most widespread use as they allow a high degree of bodily movements of the subjects. The experimental setups often utilize rhythmic interactions and use synchronization measures to characterize the interactions between agents. Using fNIRS, the study by Su et al. investigated the synchronization of behavior and the relation to cortical activity in children and adults. They report differential activation of parietal and temporal regions during observation, execution, and joint synchronized actions. In adults, compared to children, they observed a shift of activation toward leftwards lateralization in the active conditions. Maye et al. utilized EEG recordings in a full hyperscanning setup. Importantly, they used an experimental paradigm that did not involve any external rhythmic stimulation. Under these conditions, despite quite some effort, they could not observe any direct synchronization of the two brains' activity. However, they could demonstrate brain activity in different frequency bands that correlates with objective task performance, as well as the subjective rating of task performance and collaboration of the agents. Experimental investigations are not limited to the brain's activity but also include peripheral physiological signals like heart rate variability, respiration, and electrodermal activity. Maye et al. report the surprising result that the subjective evaluation of performance by the participants can be better predicted based on such autonomic parameters compared to objective behavioral parameters. This prediction advantage of the autonomous parameters disappears in the individual settings. The coupling of autonomous parameters may be modulated by the type of relationship of the two interacting agents. Balconi and Fronda present data on the modulation of heart rate and skin conductance by exchanging gifts between two agents. Specifically, the synchronization of the heart rate increased by this experimental manipulation. These studies demonstrate that investigation of the neuronal and autonomic signatures is a fruitful field for further studies.

MODELS OF SOCIAL INTERACTION

Five articles of this Research Topic address models of social interaction, ranging from basic research on understanding and modeling of the social behavior to experimental studies on deploying and testing existing models to study human-robot interaction and coordination. Tognoli et al. investigated whether the multitudinous processes associated with social behavior abide to general principles by testing a scientific approach that tightly interweaves experimental neurobehavioral studies and mathematical models. Using the Haken-Kelso-Bunz model, its application is demonstrated in the context of social coordination in several scenarios showing, for example, that accommodating for symmetry breaking in intrinsic dynamics and coupling, multiscale generalization and adaptation are principal evolutions. Bütepage et al. studied action coordination between humans and robots, in a setting where a robot is required to learn interactive tasks from a combination of observational and kinesthetic teaching. The authors demonstrate experimentally the importance of predictive and adaptive components as well as low-level abstractions to successfully learn to imitate human behavior in interactive social tasks. Demirel et al. performed a computational analysis of sensorimotor interactions in a dual-arm robotic setup, showing that, under the common fate principle, a correlation analysis of the velocities of visual pivots is sufficient to characterize "the self" (including proximo-distal arm-joint dependencies) and to assess motor to sensory influences, and "the other" by computing clusters in the correlation dependency graph. They further show that a simple correlational analysis is not sufficient to assess the non-symmetric/directed dependencies required to infer autonomy, i.e., the ability of entities to move by themselves. Maniadakis et al. studied the temporal aspects of symbiotic human-robot interaction and explore the integration of three time-aware modules to encode past and ongoing experiences, as well as the accomplishment of goals. The integrated system is then employed to coordinate the activities of a multi-agent team. Blancas et al. investigated how impairments in prediction in young adults with autism spectrum disorder relate to their behavior during collaboration. They develop a task where participants play in interaction with a synthetic agent, and the agent's behavior changes during the game, requiring adaptation and collaboration. The results show differences between autistic and neurotypical individuals in their behavioral adaptation to the other partner but no differences in the self-reports of that collaboration.

AUTHOR CONTRIBUTIONS

All authors contributed to the writing of this editorial. All authors approved the submitted version.

ACKNOWLEDGMENTS

The authors acknowledge support by the EU through project socSMCs (H2020-641321).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Psychology

ORIGINAL RESEARCH
published: 21 August 2019
doi: 10.3389/fpsyg.2019.01913

Inferring Interactivity From Gaze Patterns During Triadic Person-Object-Agent Interactions

Mathis Jording[1,2]*, Arne Hartz[3,4], Gary Bente[5], Martin Schulte-Rüther[1,3,4] and Kai Vogeley[1,2]

[1] Cognitive Neuroscience, Institute of Neuroscience and Medicine (INM-3), Research Center Jülich, Jülich, Germany, [2] Department of Psychiatry, University Hospital Cologne, Cologne, Germany, [3] JARA-BRAIN, Aachen, Germany, [4] Translational Brain Research in Psychiatry and Neurology, Department of Child and Adolescent Psychiatry, Psychosomatics, and Psychotherapy, RWTH Aachen University, Aachen, Germany, [5] Department of Communication, Michigan State University, East Lansing, MI, United States

Observing others' gaze informs us about relevant matters in the environment. Humans' sensitivity to gaze cues and our ability to use this information to focus our own attention is crucial to learning, social coordination, and survival. Gaze can also be a deliberate social signal which captures and directs the gaze of others toward an object of interest. In the current study, we investigated whether the intention to actively communicate one's own attentional focus can be inferred from the dynamics of gaze alone. We used a triadic gaze interaction paradigm based on the recently proposed classification of attentional states and respective gaze patterns in person-object-person interactions, the so-called "social gaze space (SGS)." Twenty-eight participants interacted with a computer controlled virtual agent while they assumed to interact with a real human. During the experiment, the virtual agent engaged in various gaze patterns which were determined by the agent's attentional communicative state, as described by the concept of SGS. After each interaction, participants were asked to judge whether the other person was trying to deliberately interact with them. Results show that participants were able to infer the communicative intention solely from the agent's gaze behavior. The results substantiate claims about the pivotal role of gaze in social coordination and relationship formation. Our results further reveal that social expectations are reflected in differential responses to the displayed gaze patterns and may be crucial for impression formation during gaze-based interaction. To the best of our knowledge, this is the first study to document the experience of interactivity in continuous and contingent triadic gaze interactions.

Keywords: social gaze, joint attention, eye contact, triadic interaction, non-verbal communication, social psychology, human-agent interaction

Edited by:
Andreas K. Engel,
University Medical Center Hamburg-Eppendorf, Germany

Reviewed by:
Jari K. Hietanen,
University of Tampere, Finland
Jean Baratgin,
Université Paris 8, France

***Correspondence:**
Mathis Jording
m.jording@fz-juelich.de

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: 23 April 2019
Accepted: 05 August 2019
Published: 21 August 2019

Citation:
Jording M, Hartz A, Bente G, Schulte-Rüther M and Vogeley K (2019) Inferring Interactivity From Gaze Patterns During Triadic Person-Object-Agent Interactions. Front. Psychol. 10:1913. doi: 10.3389/fpsyg.2019.01913

INTRODUCTION

During social interactions, we consistently focus on the eyes of our interaction partner because it is the fastest and easiest way to access the inner experience of another person (Yarbus, 1967; Baron-Cohen et al., 1997; Emery, 2000). From the eye region alone we are able to infer age, gender, and personality and even identify individual persons (George and Conty, 2008; Itier and Batty, 2009).

We also use gaze to ensure successful communication and smooth interactions by coordinating turn-taking (Argyle and Cook, 1976) and coordinating attention with others. This ability may constitute the phylogenetic and ontogenetic basis of cooperation (Tomasello et al., 2007; Grossmann, 2017). The most prevalent example of coordinated gaze is joint attention i.e., the joint focus of two persons gaze on an object, including gaze following and leading the gaze of others (Emery, 2000). The ability to follow someone else's gaze toward objects is acquired very early in life, possible starting at the age of 6 months (Senju and Csibra, 2008), it provides the basis for reinforcement learning (Vernetti et al., 2017), and the development of a theory of mind and language (Morales et al., 1998). It is therefore not surprising that the proficiency in gaze following predicts social competence, self-regulation abilities, and even the depth of information processing and IQ (Mundy and Newell, 2007).

During everyday encounters with other people, we do not know in advance whether the person we meet is trying to engage us in an interaction or is merely exploring the environment. In other words, we have to disambiguate the dual function of social gaze (Gobel et al., 2015; Jarick and Kingstone, 2015), or the simultaneous use of gaze for visual perception and for communicating with others. That is, we take the communicative states of others into account and adjust our gaze behavior for social adequacy accordingly (Risko and Kingstone, 2011; Wu et al., 2013). Conversely, this also implies that by observation alone we cannot be sure of whether gaze behavior of others is a communicative signal toward us or merely serves perceptual means. One powerful communicative signal is mutual eye contact (Senju and Johnson, 2009) which increases emotional empathy and modulates attention (Farroni et al., 2002; Senju and Hasegawa, 2005; Dalmaso et al., 2017). Thus, eye contact likely fosters the experience of a connection with another person. Furthermore, attempts to establish joint attention can be considered as prototypical gaze-based interaction. However, as of yet it is unclear, which cues are most informative in disambiguating the dual function of social gaze and inferring social communicative intent based on observed gaze alone.

Here we investigate the human ability to recognize communicative attempts from gaze. Using gaze-contingent paradigms with virtual characters (VC) it is possible to investigate ongoing interactions while retaining full experimental control (Vogeley and Bente, 2010; Wilms et al., 2010; Pfeiffer et al., 2013b; Georgescu et al., 2014; Oberwelland et al., 2016, 2017). However, these paradigms suffer from two major limitations: (1) gaze communication is implemented as a series of short, discrete and isolated events and not as an ongoing flux of interaction; (2) the respective paradigms mostly relied on explicit instructions or repetitive, monotonic, and predictable agent behavior. Resolving these limitations required both a theoretical foundation and technological advancements. Theoretically, we developed a new holistic taxonomy of social gaze, the "social gaze space (SGS)" (Jording et al., 2018). The SGS covers all possible categorical states of attention and interaction during gaze-based triadic interactions (constituted by two interactants and at least one object in a shared environment). The different gaze states include: "partner-oriented (PO)," during which the attention is directed solely on the interaction partner; "object-oriented (OO)," attention directed solely on the object(s) in the environment; "introspective (INT)," attention disengaged from the outside world and directed toward inner (e.g., bodily) experiences; "responding joint attention (RJA)," a state of actively following the partner's gaze toward objects of his choice; and "initiating joint attention (IJA)," a state in which the partner's gaze is led toward the objects of one's own choice. The two joint attention states (RJA and IJA) are interactive states in which the agents' behavior depends on the interaction partner, whereas the other three describe states of passive observation. Note, that these five states individually describe the behavior of one of the interaction partners. The interaction between both can be characterized as the combination of both individual states toward a "dual state" (Jording et al., 2018).

Technically, we implemented all five different gaze states of the SGS in the gaze-contingent agent-platform "TriPy" (Hartz et al., submitted). Unlike previous agent-systems, it can generate all SGS states including their responsive properties in real-time. The agent allows for mutual interactions in a continuous and immersive, hence, ecologically valid fashion. The agent's behavior is governed by sets of probabilistic parameters and timing parameters, based on empirical observations during continuous gaze-based interactions (Hartz et al., submitted).

We used this setup to address the question whether and how humans identify communicative intentions from gaze alone. To this end, we asked participants to interact with an algorithmically controlled VC while believing that a real human controlled the VC. Participants had to rate, whether their interaction partner was trying to interact with them or not. We analyzed the participants' decisions and response times (RT) as well as their gaze behavior and the occurrence of eye contact and instances of joint attention. We were interested whether participants would experience differences in the degree of interactivity of the different gaze states as implied by the SGS. We assumed that from the non-interactive states, PO would be rated the most interactive because here the agent focused on the perceiver proportionately more, increasing the probability of eye contact. With respect to interactive states, we hypothesized that the IJA state might be experienced less frequently as interactive compared to RJA. While in IJA participants need to actively follow the agent in order to learn, whether this would move the agent to "show" them the next object, in RJA the agent would strictly follow the participant which we assumed to be easily noticeable. After the experiment, we let participants rate the difficulty of the task and compared it to their performance in identifying interactive situations as an indicator of the conscious accessibility of the underlying cognitive processes.

MATERIALS AND METHODS

Participants

A total of 28 participants without any record of psychiatric or neurological illnesses were recruited via mailing lists, gave their written consent and were compensated for their participation (10€ per hour). Three participants were excluded due to technical failure (*n* = 1) and lack of conviction to interact with a real person (*n* = 2). Data from 25 participants (aged 19 – 57; mean = 31.08, *SD* = 11.21; 16 identifying as female, 9 as male) were further analyzed. This study was approved by the ethics committee of the Medical Faculty of the University of Cologne, Germany, and strictly adhered to the Declaration of Helsinki and the Principles of Good Scientific Practice.

Procedure and Tasks

Before the experiment, participants were briefly introduced to a confederate of the same sex but were brought to another room where they received the detailed written experimental instructions that were repeated orally. Participants were told that both communication partners would be represented by the same standard male VC serving as avatar and that both could only communicate via gaze behavior. They were further told that they would be seated in front of a monitor that displayed the avatar of their partner representing the partner's eye movements on the basis of data provided by two identical eye-tracking systems and updating the respective gaze direction of the avatars in real-time (**Figure 1A**). In fact, participants always solely communicated with an agent controlled by a computer algorithm (Hartz et al., submitted). Participants would further see four trial-wise changing objects, at fixed positions and obviously visible for the partner's avatar (**Figure 1B**). Neither the VC nor the objects were shown to the participant before the start of the experiment.

Participants were further instructed to take two different roles: (1) The Observation-Role (ObR), and (2) the Action-Role (AcR). For the ObR condition, there were no trial specific instructions apart from the task to ascertain whether their partner was trying to "interact" or not (German "austauschen" or "interagieren"), "interacting" was defined as an encounter in which both partners respond to the gaze behavior of the partner in a mutual and reciprocal fashion. Participants were asked to answer only as soon as they felt "quite sure" but were reminded that each trial ended at the latest after 30 s and they therefore would have to hurry. The time between beginning of the trial and button press was logged as RT. When participants had not pressed a button within 30 s, they were asked to decide more quickly in the next trial. After each trial, the participant's choice was displayed on the screen until participants indicated their readiness to continue via button press. Afterward, a message was displayed, asking the participants to wait until their partner was ready for the next trial. This delay was introduced in order to support the participants believe in the confederate based coverstory. The next trial would then begin after a random (uniformly distributed) duration of 1 – 5 s with the appearing of the agents face on the screen.

During the AcR condition, participants were explicitly instructed to engage in one of the states of the SGS (Jording et al., 2018) with the following instructions: "Please concentrate on your partner" (German: "Bitte konzentrieren Sie sich auf Ihren Partner"; PO); "Please attend to the objects" (German: "Bitte achten Sie auf die Objekte"; OO); "Please keep your eyes open and concentrate on your breath" (German: "Bitte lassen Sie Ihre Augen geöffnet und konzentrieren Sie sich auf Ihren Atem"; INT); "Please interact with your partner and let his gaze guide you" (German: "Bitte versuchen Sie sich mit Ihrem Partner auszutauschen und lassen Sie sich von seinem Blick leiten"; RJA), or "Please interact with your partner and use your gaze to guide him" (German: "Bitte versuchen Sie sich mit Ihrem Partner auszutauschen und nutzen Sie Ihren Blick um ihn zu leiten"; IJA). No further instructions were given and participants were told that there was no correct or wrong behavior and they should behave according to their intuitive understanding of these instructions. Trials stopped after 30 s and were followed by a short break of 2 – 6 s.

Whereas ObR was the target condition allowing measuring the experience of interactivity, the AcR condition was included to support the cover story, as participants believed to be interacting with some other real participants and thus would expect a balanced study design with the same tasks for both participants. Both roles were presented alternatingly in three blocks each, with 16 trials per block during ObR and 10 trials per block for AcR. The order of blocks and state instructions within blocks was randomized across participants. After two blocks participants were given a short break of up to 3 min to prevent fatigue and to allow for recalibration of the eyetracker to avoid drifting artifacts.

Setup, Agent-Platform, and Pilot Study

The setup consisted of an eye-tracker with a sampling rate of 120 Hz and an accuracy of 0.5° (Tobii TX300; Tobii Technology, Stockholm, Sweden). A 23" monitor with a screen resolution of 1920*1080 pixels mounted on top of the eye-tracker was used as display (**Figure 1A**). Participants were seated at a distance between 50 – 70 cm to the monitor. A PC-keyboard with the marked buttons "J" and "N" was used for participant responses during ObR. A light sensor based system (StimTracker, Cedrus Corporation, San Pedro, CA, United States) ensured that timing of presented stimuli by the algorithm and actual graphical output were in sync.

The agent's behavior and graphical output was controlled by the agent-platform "TriPy" (Hartz et al., submitted), implemented in Python 2.7 (Python Software Foundation[1]) using PyGaze (Dalmaijer et al., 2014) and PsychoPy (Peirce, 2008). TriPy is based on a gaze-contingent algorithm that adapts the behavior of a VC to the behavior of the participant in real-time (Wilms et al., 2010). In contrast to previous setups, TriPy does not require a prior determination of the exact course and timing of the agents' behavior. Instead, behavior in the non-interactive states is implemented on a probabilistic basis in which the agent displays different micro states (e.g., a moment of looking at one of the objects) with different probabilities (**Figure 2**). In the RJA state the agent follows the participants gaze toward the objects and looks back at the eyes of the participant, when being looked at himself, with a randomly drawn offset between 311.06 – 589.93 ms (lognorm distributed, range 6.06 ± 0.32). In the IJA state the agent looks at the participant and as soon as eye contact is established or after a randomly drawn waiting period of 772.78 – 2321.57 ms (lognorm distributed, range 7.2 ± 0.55) looks at one of the objects at random. As soon as the participant follows or after a randomly drawn waiting period of 780.55 – 2440.60 ms (lognorm distributed, range

---

[1] https://www.python.org

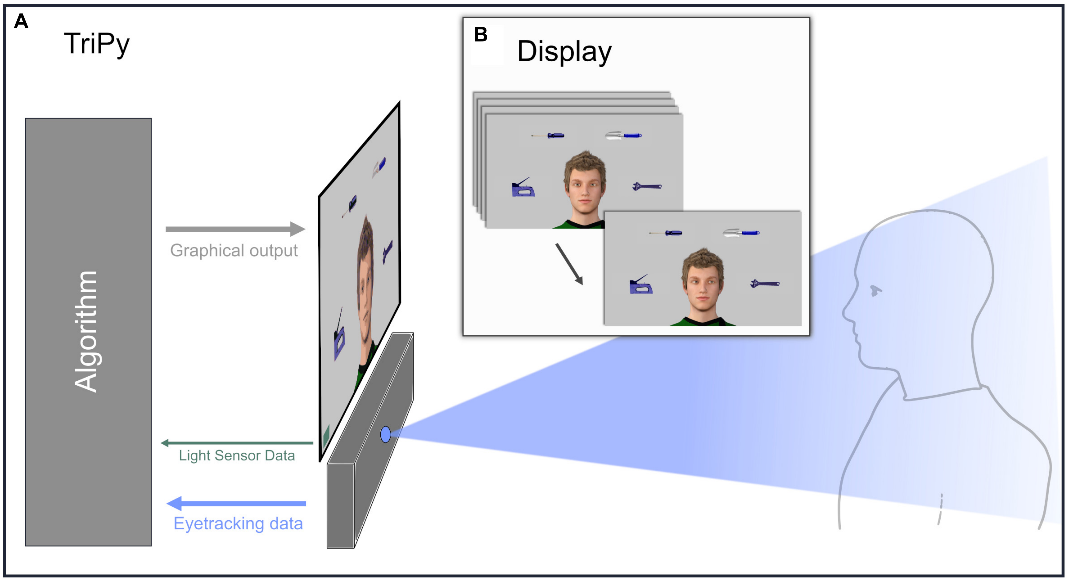


FIGURE 1 | Illustration of the technical setup and the participants' perspective during the experiment. **(A)** Illustration of a participant interacting with the agent controlled by the platform TriPy. **(B)** The behavior of the agent created by TriPy as seen from the perspective of the participant **(B)**.

7.23 ± 0.57), the agent starts anew with trying to establish eye contact and subsequently choosing a new object at random (video examples of the agents behavior in all states can be found in the **Supplementary Material**). These microstates, their durations and transition probabilities, as well as temporal parameters of the interactive agents' states were empirically informed by a pilot study (Hartz et al., submitted). The anthropomorphic VC was created with the modeling software Daz Studio 3.1 (DAZ Productions, Inc., United States).

During the ObR condition, the agent equally often displayed either any of the interactive (25% for each of the interactive states RJA and IJA) or any of the non-interactive states (16.67% for each of the non-interactive states PO, OO, and INT). This partitioning ensured that participants encountered interactive and non-interactive states equally often and thus could not exceed a 50% correctness rate by guessing. During AcR – which was established only to let participants continuously believe that they were interacting with the interaction partner to whom they had been introduced before the experiment – the agents' states corresponded to the states of the participant the agent displayed non-interactive states (PO, OO, or INT) when the participant herself was in a non-interactive state with all combinations of agent and participant states appearing equally often. Each interactive-state of the participant was answered by the agent with the complementary interactive-state (RJA with IJA; IJA with RJA).

Questionnaires and Post-experimental Inquiry and Information

After the experiment participants filled out a post-experimental questionnaire asking on visual analog scales (ranging from 1 to 6): (A1) how difficult they had experienced the ObR tasks, (A2) how difficult the AcR tasks, (A3) how natural they had experienced the interaction, and (A4) how they rated the quality of the technical realization of the VC's eye movements. In addition, participants were given the chance to respond in open texts relating to: (B1) their assumptions as to the purpose of the study, (B2) anything that bothered them during tasks of both types ObR and AcR, (B3) any strategies they had employed in their attempt to communicate with the other person, (B4) how the naturalness of the interaction could be improved, (B5) whether there was anything else to the experiment which bothered them. The participants' belief in the cover story was further tested in an interview by the experimenter. Participants were asked how well the communication with the partner had worked, whether they had considered what their partner was thinking and whether they had tried to empathize with their partner and whether they had applied specific strategies in their communication with the partner. In addition to the post-experimental questionnaire, participants, either before or after the experiment, also answered a demographic questionnaire and the German version of the autism-spectrum-quotient (AQ; Baron-Cohen et al., 2001). However, for none of the participants AQ results pointed toward autistic symptomatology (cut-off > 32; Baron-Cohen et al., 2001). After the experiment, interview, and questionnaires participants were informed about the nature of the cover story and explained its necessity. Now, participants were asked directly, whether they have had any suspicions as to the nature of the experiment or their partner.

Data Preprocessing and Statistical Analysis

From a total 1200 trials in the ObR condition (25 participants with 48 trials each), 39 trials were excluded due to missing

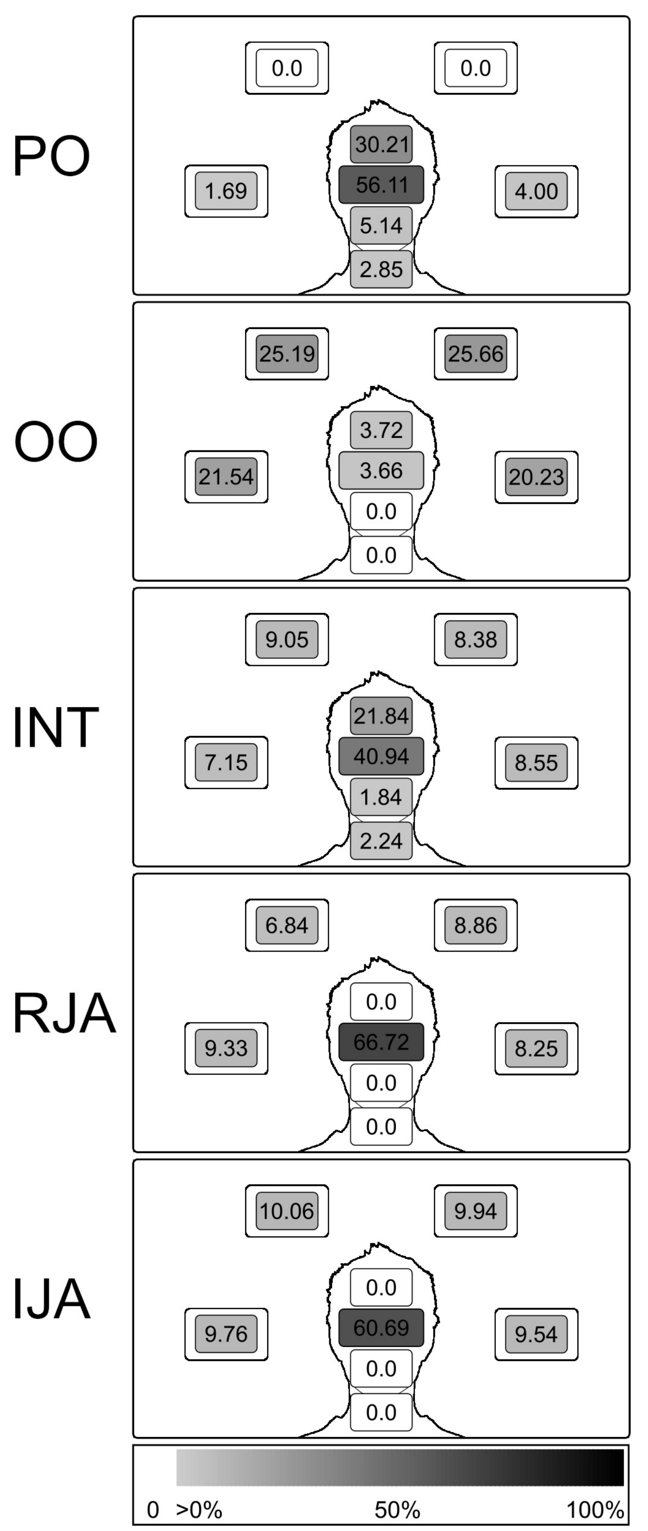


FIGURE 2 | Illustration of the distribution of the agent's visual attention separately cumulated for the different gaze states. Numbers express the rate in percent with which the agent looked at the AoIs in total in the specific state as portion of all fixations, color schemes coding serve as additional illustration (white, AoIs not being targeted; light gray, low rate; black, high rate; see color bar legend at the bottom).

responses or RT exceeding 30 s, another 201 trials were excluded because more than 20% of gaze data were missing due to technical problems, 960 trials remained for analysis. Response, eye-tracking, and questionnaire data were preprocessed and statistically analyzed with R (R Development Core Team, 2008) and RStudio (RStudio Team, 2016). Response and eye-tracking data were analyzed with (generalized) linear mixed effects models, as recommended for data from repeated measures designs (Pinheiro and Bates, 2009), using the lmer() and glmer() function from the lme4 package (Bates et al., 2015). The general influence of predictors was assessed in likelihood ratio tests, comparing how well models including different predictors fit a given data set while taking into account (i.e., penalizing) the models' complexity. The significance of the effect of each predictor was tested by comparing a model comprising the predictor with the same model without the predictor against a significance level of 0.05. Where likelihood ratio tests revealed significant effects of factors, we conducted Tukey *post hoc* tests for the comparison between all individual factor levels (correcting for multiple comparisons) with the glht() function from the multcomp package (Hothorn et al., 2008).

For the analysis of gaze data we computed "relative fixation durations" as the portion of cumulative fixation durations spent on the AoIs "eyes", "face" (not including the eyes), or "objects" (the four objects taken together). Instances of eye contact and joint attention were defined as situations in which the participant and the agent both looked at the eyes of the partner (eye contact) or simultaneously at the same object (joint attention). Two consecutive eye contact or joint attention events on the same object were treated as a single continuous event when they were less than 100 ms apart in order to prevent artificial inflation of events due to eye blinks. Only eye contact and joint attention events with a minimum duration of 50 ms were included in the analysis.

Data from the visual analog scales in the post-experimental questionnaire were summarized as group means. In addition, Spearman correlations between participants' post-experimental self-reports and their task performance were computed. The effect of the participants' age and gender on their responses were analyzed in linear models. Open text responses and statements from the interview were checked for any indications of mistrust in the cover story (e.g., statements indicating lack of conviction to interact with a real person).

RESULTS

Interactivity Ratings

In order to test whether participants were able to correctly identify interactive situations we first compared within ObR the ratings between the non-interactive states (PO, OO, and INT) and the interactive states (RJA and IJA) as a logistic regression with random intercepts for participants. The analysis revealed a highly significant effect on the model fit [$\chi^2(1) = 222.59$, $p < 0.001$]. The chance of being rated as interactive was 27.07% for the non-interactive states and 73.32% for the interactive states, corresponding to a difference in the predicted odds ratio by the factor of 8.45 ($M = 2.13$, $SD = 0.16$).

In a next step we looked at the difference between the individual states (**Figure 3A**), again analyzed as logistic regression with random intercepts for participants. A model comprising the agent state as fixed effects fitted the data significantly better than the null model including only the intercept [$\chi^2(4) = 266.70$, $p < 0.001$]. *Post hoc* tests revealed significantly lower ratings for PO vs. INT ($M = -0.86$, $SD = 0.26$, $z = -3.30$, $p = 0.009$), INT vs. RJA ($M = -1.06$, $SD = 0.22$, $z = -4.79$, $p < 0.001$), and RJA vs. IJA ($M = -1.13$, $SD = 0.23$, $z = -4.92$, $p < 0.001$), but not between OO and PO ($M = -0.17$, $SD = 0.28$, $z = -0.60$, $p = 0.975$). Note that for the sake of simplicity we only report comparisons between neighboring ranks when sorted by mean estimates. All other comparisons between states yielded highly significant differences (all $p < 0.001$).

RTs (**Figure 3B**), were logarithmized and again analyzed in a linear mixed effects model with random intercepts for subjects. A group-wise comparison between the interactive and the non-interactive states as fixed effects had no significant effect on the model fit [$\chi^2(1) = 0.36$, $p < 0.55$]. However, including the individual agent states in the model as fixed effects proofed to fit the data significantly better than the null model [$\chi^2(4) = 82.55$, $p < 0.001$]. Corresponding to the results from the interactivity ratings, *post hoc* tests revealed significant differences between OO & PO ($M = -0.18$, $SD = 0.04$, $z = -4.49$, $p < 0.001$), PO & INT ($M = -0.12$, $SD = 0.04$, $z = -2.85$, $p = 0.035$), INT & RJA ($M = 0.22$, $SD = 0.04$, $z = 5.83$, $p < 0.001$), and RJA & IJA ($M = -1.84$, $SD = 0.03$, $z = -5.55$, $p < 0.001$). Note that the differences between OO & INT ($M = -0.30$, $SD = 0.04$, $z = -7.33$, $p < 0.001$), PO & RJA ($M = 0.10$, $SD = 0.04$, $z = 2.748$, $p = 0.048$), and OO and IJA ($M = -0.26$, $SD = 0.04$, $z = -7.17$, $p < 0.001$) also reached significance. In order to investigate whether the quality of the participants' ratings would increase with longer decision time we computed mean correctness scores (RC; correct = "non-interactive" for PO, OO, and INT or "interactive" for RJA and IJA) for each participant. We found a significant relationship between the participants' mean RC and mean RT (r = 0.45, $p < 0.05$). In addition, we analyzed, whether the participants' age or gender had an influence on their decisions. However, neither age nor gender had any significant effect on the mean RCs [age: $\chi^2(1) = 2.21$, $p < 0.151$; gender: $\chi^2(1) = 2.12$, $p < 0.159$] or mean RTs [age: $\chi^2(1) = 0.518$, $p < 0.479$; gender: $\chi^2(1) = 1.43$, $p < 0.245$].

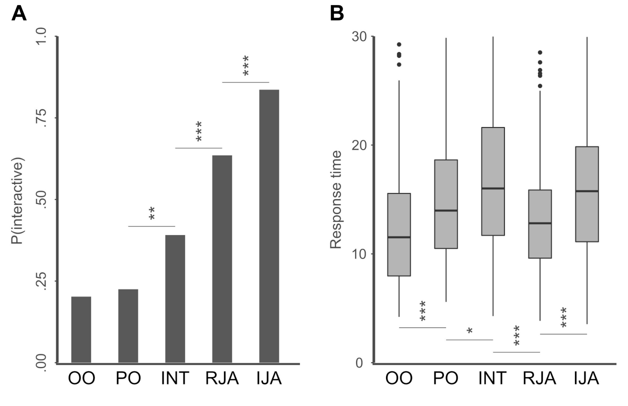


FIGURE 3 | Plots of mean interactivity ratings and mean response times separately for the different gaze states. **(A)** Mean interactivity ratings for different agent states. Asterisks indicate significant differences between neighboring states (when ranked in ascending order) in *post hoc* tests (* <0.05; ** <0.01; and *** <0.001). **(B)** Mean RTs in ms for different agent states. Asterisks indicate significant differences between neighboring states (when ranked in ascending order of mean interactivity ratings) in *post hoc* tests (* <0.05; ** <0.01; and *** <0.001).

Gaze Behavior

For the participants' gaze behavior during ObR, we analyzed the effect of non-interactive vs. interactive states, of the AoIs Eyes, Face and Object and the interaction between states and AoIs on relative durations (proportion of cumulative fixation durations from 0 to 1, **Figure 4A**). Tests did not reveal significant improvements in model fit for including states [$\chi^2(1) = 0.00$, $p = 0.994$] but for AoI [$\chi^2(2) = 948.37$, $p < 0.001$], and the interaction of state*AoI [$\chi^2(2) = 12.40$, $p = 0.002$]. A *post hoc* test between factor combinations was conducted in order to identify effects potentially driving the interaction. However, corrected for multiple testing, the comparisons between non-interactive and interactive states did not reveal any significant differences for the AoIs Eyes ($M = -0.03$, $SD = 0.02$, $z = -1.80$, $p = 0.467$), Face ($M = -0.03$, $SD = 0.02$, $z = -1.64$, $p = 0.565$), or Objects ($M = -0.04$, $SD = 0.07$, $z = -2.58$, $p = 0.102$).

The effect of a non-interactive vs. interactive agent on the number of instances of eye contact (**Figure 4B**) and joint attention (**Figure 4C**) per trial was analyzed in generalized mixed effects models for Poisson distributed data. Including the interactivity of the agent significantly increased model fits for the prediction of the amount of eye contact [$\chi^2(1) = 68.19$, $p < 0.001$] as well as the amount of joint attention instances [$\chi^2(1) = 72.75$, $p < 0.001$]. When the agent behaved interactively, the occurrence of eye contact instances increased by a factor of 1.31 ($M = 0.27$, $SD = 0.03$) and the occurrence of joint attention instances increased by a factor of 1.52 ($M = 0.42$, $SD = 0.05$).

We then analyzed whether the occurrence of instances of eye contact (**Figure 4D**) or joint attention (**Figure 4E**) had a predictive value for the participants' subsequent interactivity rating and whether the prediction would differ depending on the agent behaving either non-interactively or interactively. To this end, we compared linear mixed effects models including the agents' interactivity, the number of instances of eye contact or joint attention, respectively, as well as the interaction between both. All three, the inclusion of the agents' interactivity [$\chi^2(1) = 222.57$, $p < 0.001$], the inclusion of the number of eye contact instances [$\chi^2(1) = 14.86$, $p < 0.001$], as well as the interaction between both [$\chi^2(1) = 9.52$, $p = 0.002$], and significantly improved model fits. The predicted probability of the agents′ behavior being rated as interactive increased with the number of eye contact instances ($M = 0.05$, $SD = 0.03$), but this effect was especially strong when the agent actually behaved interactively ($M = 0.15$, $SD = 0.05$). For the analysis of the effect

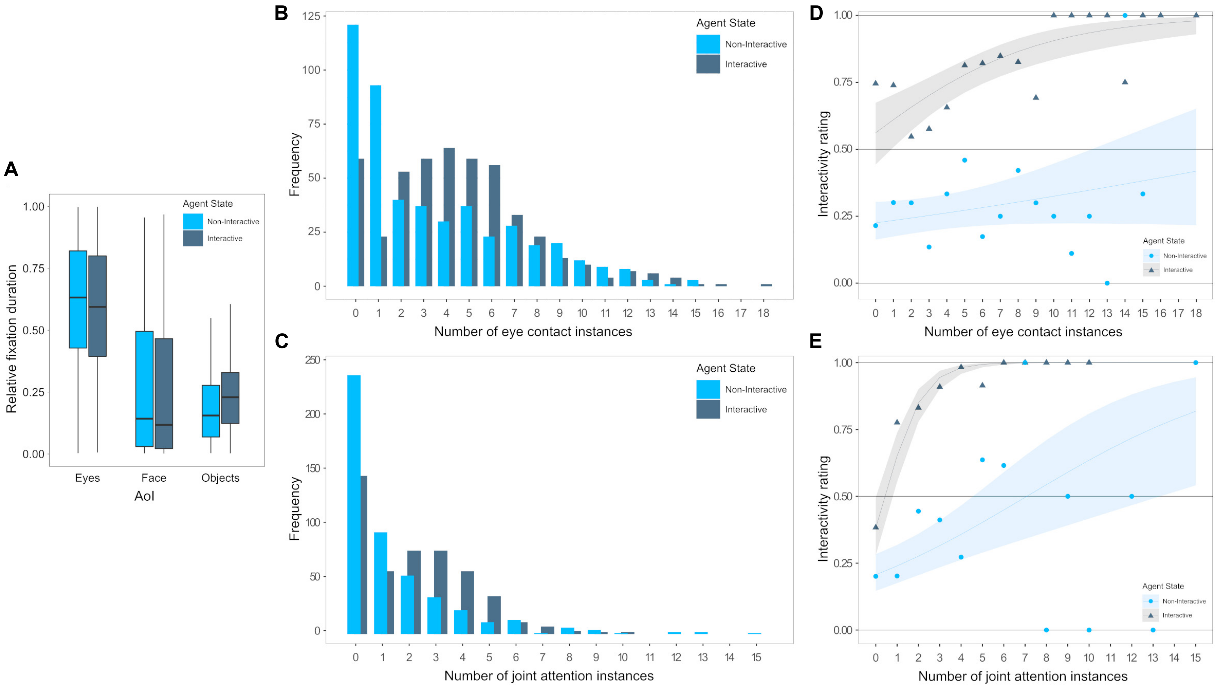


FIGURE 4 | Illustration of the participants gaze behavior and instances of eye contact and joint attention between participant and agent in connection to the participant's rating of the agents interactivity, separately for an agent behaving non-interactively (light blue) vs. interactively (dark blue). **(A)** Boxplots of relative fixation durations as the portion of time spent on the AoIs Eyes, face, and objects per trial. **(B)** Frequencies of eye contact instances per trial. **(C)** Mean rates (circles and triangles) and model predictions with 95% confidence intervals (lines and ribbons) of interactivity ratings for differing numbers of eye contact instances per trial. **(D)** Frequencies of joint attention instances per trial. **(E)** Mean rates (circles and triangles) and model predictions with 95% confidence intervals (lines and ribbons) of interactivity ratings for differing numbers of joint attention instances per trial.

of joint attention, again, the inclusion of the agents' interactivity [$\chi^2(1) = 222.59$, $p < 0.001$], the inclusion of the number of joint attention instances [$\chi^2(1) = 96.54$, $p < 0.001$], as well as the interaction between both [$\chi^2(1) = 73.16$, $p < 0.001$], significantly improved model fits. Accordingly, the predicted probability of the agents' behavior being rated interactive increased with the number of joint attention instances ($M = 0.19$, $SD = 0.05$) with an even stronger effect when the agent actually behaved interactively ($M = 0.92$, $SD = 0.12$).

Questionnaires and Post-experimental Inquiry

In the post experimental inquiry participants reported on the perceived difficulty of the ObR task ($M = 2.80$, $SD = 1.38$) and the AcR task ($M = 1.76$, $SD = 0.72$), the quality of the technical implementation of the agents' eye movements ($M = 3.21$, $SD = 0.88$), and the naturalness of the interaction ($M = 2.96$, $SD = 1.30$). We compared ratings of the task difficulty to the participants' mean tendency to experience the agent as interactive, their mean performance (response correctness) as well as mean RTs. Difficulty ratings neither correlated significantly with the participants' tendency to rate the agent's behavior as interactive ($rs = -0.07$, $p > 0.05$) nor with their response correctness ($rs = 0.02$, $p > 0.05$) nor with RTs ($rs = -0.24$, $p > 0.05$).

In order to assess effects of autistic traits we compared models comprising and not comprising the AQ scores as predictor. Neither including the quotient as main effect [$\chi^2(1) = 0.98$, $p < 0.323$] nor as interaction with interactive vs. non-interactive states [$\chi^2(1) = 0.27$, $p < 0.607$] significantly improved model fits for mean interactivity ratings. Similarly, for mean RTs, neither including the quotient as main effect [$\chi^2(1) = 0.45$, $p < 0.50$] nor as interaction with interactive vs. non-interactive states [$\chi^2(1) = 0.01$, $p < 0.908$] significantly improved model fits.

None of the answers to the written open text questions indicated any suspicions about the cover story or any awareness of deceit. In the interview, two participants indicated that during the experiment they developed the suspicion or had asked themselves whether they actually had interacted with the partner they previously had met (both participants were excluded from further analysis, see above).

DISCUSSION

This study focuses on the question whether and how humans are able to recognize interactivity in triadic interactions. To this extent, we gave our participants two tasks, one in which participants had to observe and recognize gaze states (ObR) and one in which they had to engage in different gaze states (AcR). While the former condition was the actual target condition and basis for the analysis, the latter was necessary to maintain the semblance of a balanced study design suggested by the cover story. As our main result, we can show for the first time that human participants are perfectly able to use gaze cues to judge interactivity by spotting the contingencies between their own and the agents' behavior without any explicit instructions how to do that. In the analysis of the interactivity ratings, we found that participants consistently and successfully discriminated between interactive and non-interactive states. These findings empirically substantiate the hypothesis of gaze communication being a precursor of human cooperation (Moll and Tomasello, 2007; Tomasello et al., 2007). Findings from phylogenetic and ontogenetic studies support this notion by showing that attending to eyes and communicating via gaze are pivotal steps toward higher levels of social cognition (Tomasello and Carpenter, 2005; Tomasello et al., 2007; Grossmann, 2017). So far, however, these proposals have been hypothetical, i.e., based on phylogenetic and evolutionary considerations. Here, we can explicitly show that gaze is sufficient for humans to establish the experience of mutual interaction as a prerequisite for building social relationships.

We also found differences in the interactivity ratings within interactive-states and within non-interactive states suggesting considerable sensitivity to variations in the tempo-spatial parameters of perceived gaze behavior. Our expectation that a gaze following agent would more easily elicit the experience of interactivity was not confirmed. This hypothesis was based on the assumption that actively following an initiating agent would be more demanding than being followed by a responding agent. Earlier studies had shown that humans innately expect gaze following (Pfeiffer et al., 2011) and perceive the initiation of joint attention as rewarding (Schilbach et al., 2010; Pfeiffer et al., 2014; Oberwelland et al., 2016). However, the present data suggest that agents who initiate joint attention are significantly more readily experienced as interactive than a merely gaze following agent. This might be explained by the fact that responding to joint attention bids might be considerably easier than to actively initiate joint attention. This interpretation is in accordance with phylogenetic and ontogenetic findings suggesting that IJA requires more complex cognition as compared to RJA. For example, chimpanzees are able to follow someone's gaze but do not initiate joint attention themselves (Tomasello and Carpenter, 2005). Human children acquire the basis of RJA from the early age of 6 month in comparison to the initiation of attention which does not occur before the second year of life (Mundy and Newell, 2007; Mundy et al., 2007).

The non-interactive states OO and PO were significantly more often identified correctly as non-interactive than the INT state. During OO the agent was mainly focused on the objects and looked at the participant only to a lesser extent. Humans are typically very sensitive to how other persons explore and behave in a shared environment. Our perception and processing of objects seem to be fundamentally altered when we observe other person attending to them (Becchio et al., 2008). Objects subsequently appear more familiar (Reid et al., 2004; Reid and Striano, 2005) and likeable (Bayliss and Tipper, 2006; Bayliss et al., 2006). Our results suggest that despite such effects, we are still able to discern that the behavior we observe is not related to us or at least not aimed at us. The same might be true for the PO state. Contrary to our prior hypothesis, participants did not report the PO agent as more interactive than OO, notwithstanding the higher chances of eye contact in these situations due to the agent more frequently looking at the participant. The instructions defined an interaction in terms of mutual and reciprocal responses between both partners. Low interactivity ratings for PO might therefore be just a sign for the participants' adherence to the instructions instead of disclosing their intuitive, subjective definition of an interaction. Despite that, participants were able to differentiate between an active, reciprocal interaction and person-focused but passive visual attention. This is in line with findings showing that humans are very sensitive to differences in the interactional affordance in the context of more pronounced contrast between encountering real persons as compared to facing static pictures (Hietanen et al., 2008; Pönkänen et al., 2011).

In our experimental setup, INT appears to be the most ambiguous of all states, receiving almost as many interactive as non-interactive ratings. The inward directed attention and thus absence of any obvious attentional focus in the environment probably made it impossible to attribute intentions of interaction. In other words, gaze alone is no longer informative as soon as the interaction partner is in a state of introspection or mind-wandering (see section "Limitations").

In order to better understand the emergence of the experience of interaction, we analyzed the relationship between the gaze behavior of the participants and the agent's behavior. We did not find any effect of the agents' intended interactivity of the encounter on the distribution of the participants visual attention between objects and agent. However, when looking at the synchronization with the agent's behavior, we found an increase in the number of eye contact instances and joint attention instances in interactive as compared to non-interactive states. Thus, one of the participants' strategies to judge upon interactivity might have been based on the frequency of eye contact and joint attention instances. The analysis of the effect of the number of eye contact and joint attention instances on the participants' decisions revealed significant differences between non-interactive and interactive encounters. Importantly, during interactive encounters, the emergence of eye contact and joint attention had much higher effects on the subsequent interactivity ratings. One plausible interpretation could be that participants "tested" the agents' reciprocity by attempting to establish eye contact and joint attention and subsequently assessing whether the timing of resulting joint contingencies could be attributed to an interacting agent that takes into account the gaze behavior of the participant. Considering the importance of fine-grained timing during such gaze-based interactions it is plausible that the emergence of interactivity is deeply embedded in the temporal enfolding of gaze-based encounters and can only be experienced over time. This is in line with the understanding that non-verbal communication is a dynamic and continuous process (Burgoon et al., 1989) that cannot be fully comprehended through the passive observation of discrete events, uncoupled from the flow of communication.

With respect to the differences in the duration of the decision between the different conditions, we found a correlation between the mean RT of participants and mean correctness scores, suggesting that participants who invested more time were able to make better informed decisions. When comparing RT between states on a single trial level, RTs in non-interactive states showed a pattern roughly corresponding to that of the correctness scores. i.e., RTs reflected the ambiguity and associated difficulty to judge the interactivity. When comparing the participants' reactions to RJA vs. IJA agents we found longer RTs for the more unequivocal IJA state (as reflected in higher interactivity ratings). One explanation might be that participants needed more time to identify this maximal complex state.

Previous studies about social gaze, even those employing gaze-contingent interactive paradigms, were mostly based on a trial structure that sharply restricted the interaction to a few seconds (Wilms et al., 2010; Pfeiffer et al., 2011, 2012; Oberwelland et al., 2016, 2017). Our findings suggest that such short time intervals are probably not sufficient to establish the full experience of interaction during a spontaneous encounter. Earlier studies circumvented this problem by focusing on "atomic" elements of interaction using an exactly predefined time course of specific behavioral elements and explicitly instructing participants. However, this restriction is not compatible with the implicit and dynamical character of social interactions and thus threatens ecological validity (Risko et al., 2012, 2016; Pfeiffer et al., 2013a; Schilbach et al., 2013).

Overcoming this problem required both theoretically and methodologically new approaches. From a theoretical perspective the SGS provides the holistic framework that is able to encompass and describe the entire span of possible interactive states (Jording et al., 2018). Methodologically this study profits from the development of the new agent-platform TriPy that implements the states of the SGS and allows for a degree of interactional freedom not available with previous setups (Hartz et al., submitted). In combination, these developments allowed us for the first time to investigate the unfolding of a purely gaze based interaction.

Limitations

Several limitations with respect to the study design need to be considered when interpreting the results. First, we deliberately focused on gaze and restricted all communication to this particular important non-verbal communication channel. The availability of additional channels would certainly have facilitated the establishment of interactions in this study, resulting in more decisive, and faster interactivity ratings. However, the goal of this study was to test explicitly the potential of gaze communication to establish interactions in a way that results can inform studies about non-verbal multi-channel communication. Furthermore, we aimed at studying the individual characteristics of predefined states of gaze interactions and therefore chose a design where the agents displayed only one state at a time. Based on these results it would now be interesting to investigate how transitions between these states might take place (Jording et al., 2018). Therefore, sampling experiences of participants at random time points in an interaction with an agent who dynamically transitions from one state to another might constitute a promising approach.

We did not aim for the systematic investigation of effects of inter-individual differences during the establishment of gaze interactions and while we included a broad age range, we did not balance our sample with regard to gender. In addition, we only used one VC with a male, middleaged appearance and did not systematically match age and gender between participants and agent. Although we did not find any significant effects of age or gender on the quality or timing of the participants' ratings, we cannot rule out the possibility of any influence. Further investigations controlling for the participants' age and gender distribution and a systematic matching between participants and agents are required to elucidate this question.

Conclusion

Results indicate that humans are able to establish gaze interaction without any instructions or additional communication channels, supporting theoretical assumptions of the fundamental role of gaze communication in the development of human social behavior. Our data suggest that human participants are able to identify interactivity not only based on passive observation but potentially by actively studying the agents' responsiveness based on successfully established mutual eye contact and joint attention. However, participants were not only able to distinguish interactive and non-interactive situations, but behavioral differences between the non-interactive states elicited differential experiences of the interaction. Interestingly, the participants' performance did not predict their post-experimental assessment of the tasks difficulty. This suggests that decisions were based on intuition or at least partly beyond conscious processing, which corresponds to the presumably implicit and automatic character of non-verbal communication (Choi et al., 2005). An intriguing next step would now be to integrate additional non-verbal communication channels, potentially in a more immersive environment (e.g., a virtual reality), or to investigate the establishment of interactions in cases of impaired communication abilities as in autism spectrum conditions.

DATA AVAILABILITY

The datasets for this manuscript are not publicly available because the supervising ethics committee has not yet approved the publication of the raw data. Requests to access the datasets should be directed to the corresponding author.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Principles of Good Scientific Practice, ethics committee of the Medical Faculty of the University of Cologne, Germany, with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the ethics committee of the Medical Faculty of the University of Cologne, Germany.

AUTHOR CONTRIBUTIONS

All authors substantially contributed to the conception of the work. AH, MJ, KV, and MS-R designed the study protocol. AH implemented the paradigm code. MJ conducted the pilot study and the main experiment. MJ and AH analyzed the data. MJ drafted the manuscript. AH, GB, KV, and MS-R critically revised the manuscript.

FUNDING

This study was received funding from the Federal German Ministry of Education and Research (Grant Numbers: 16SV7242 to MS-R and 16SV7244 to KV), the Excellence Initiative of the German federal and state governments (MS-R), and the EC, Horizon 2020 Framework Program, FET Proactive (Grant Agreement Id: 824128 to KV).

ACKNOWLEDGMENTS

The authors want to express their gratitude to Björn Guth for his contributions to the development of the experimental code.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01913/full#supplementary-material

REFERENCES

Argyle, M., and Cook, M. (1976). *Gaze and Mutual Gaze*. Cambridge: Cambridge University Press.

Baron-Cohen, S., Wheelwright, S., and Jolliffe, T. (1997). Is there a "Language of the Eyes"? evidence from normal adults, and adults with autism or asperger syndrome. *Vis. Cogn.* 4, 311–331. doi: 10.1080/713756761

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., and Clubley, E. (2001). The autism-spectrum quotient (AQ): evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. *J. Autism Dev. Disord.* 31, 5–17.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Bayliss, A. P., Paul, M. A., Cannon, P. R., and Tipper, S. P. (2006). Gaze cuing and affective judgments of objects: i like what you look at. *Psychon. Bull. Rev.* 13, 1061–1066. doi: 10.3758/BF03213926

Bayliss, A. P., and Tipper, S. P. (2006). Predictive gaze cues and personality judgments: should eye trust you? *Psychol. Sci.* 17, 514–520. doi: 10.1111/j.1467-9280.2006.01737.x

Becchio, C., Bertone, C., and Castiello, U. (2008). How the gaze of others influences object processing. *Trends Cogn. Sci.* 12, 254–258. doi: 10.1016/j.tics.2008.04.005

Burgoon, J. K., Buller, D. B., and Woodall, W. G. (1989). *Nonverbal Communication: The Unspoken Dialogue*. New York, NY: Harper & Row.

Choi, Y. S., Gray, H. M., and Ambady, N. (2005). "The glimpsed world: unintended communication and unintended perception," in *The New Unconscious*, eds R. R. Hassin, J. S. Uleman, and J. A. Bargh (New York, NY: Oxford University Press), 309–333.

Dalmaijer, E. S., Mathôt, S., and Van der Stigchel, S. (2014). PyGaze: an open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behav. Res. Methods* 46, 913–921. doi: 10.3758/s13428-013-0422-422

Dalmaso, M., Castelli, L., Scatturin, P., and Galfano, G. (2017). Trajectories of social vision: eye contact increases saccadic curvature. *Vis. Cogn.* 25, 1–8. doi: 10.1080/13506285.2017.1290727

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neurosci. Biobehav. Rev.* 24, 581–604. doi: 10.1016/S0149-7634(00)00025-27

Farroni, T., Csibra, G., Simion, F., and Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proc. Natl. Acad. Sci. U.S.A.* 99, 9602–9605. doi: 10.1073/pnas.152159999

George, N., and Conty, L. (2008). Facing the gaze of others. *Neurophysiol. Clin. Neurophysiol.* 38, 197–207. doi: 10.1016/j.neucli.2008.03.001

Georgescu, A. L., Kuzmanovic, B., Roth, D., Bente, G., and Vogeley, K. (2014). The use of virtual characters to assess and train non-verbal communication in high-functioning autism. *Front. Hum. Neurosci.* 8:807. doi: 10.3389/fnhum.2014.00807

Gobel, M. S., Kim, H. S., and Richardson, D. C. (2015). The dual function of social gaze. *Cognition* 136, 359–364. doi: 10.1016/j.cognition.2014.11.040

Grossmann, T. (2017). The eyes as windows into other minds: an integrative perspective. *Perspect. Psychol. Sci.* 12, 107–121. doi: 10.1177/1745691616654457

Hietanen, J. K., Leppänen, J. M., Peltola, M. J., Linna-aho, K., and Ruuhiala, H. J. (2008). Seeing direct and averted gaze activates the approach–avoidance motivational brain systems. *Neuropsychologia* 46, 2423–2430. doi: 10.1016/j.neuropsychologia.2008.02.029

Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general parametric models. *Biom. J.* 50, 346–363. doi: 10.1002/bimj.200810425

Itier, R. J., and Batty, M. (2009). Neural bases of eye and gaze processing: the core of social cognition. *Neurosci. Biobehav. Rev.* 33, 843–863. doi: 10.1016/j.neubiorev.2009.02.004

Jarick, M., and Kingstone, A. (2015). The duality of gaze: eyes extract and signal social information during sustained cooperative and competitive dyadic gaze. *Front. Psychol.* 6:1423. doi: 10.3389/fpsyg.2015.01423

Jording, M., Hartz, A., Bente, G., Schulte-Rüther, M., and Vogeley, K. (2018). The "Social Gaze Space": a taxonomy for gaze-based communication in triadic interactions. *Front. Psychol.* 9:226. doi: 10.3389/fpsyg.2018.00226

Moll, H., and Tomasello, M. (2007). Cooperation and human cognition: the Vygotskian intelligence hypothesis. *Philos. Trans. R. Soc. B Biol. Sci.* 362, 639–648. doi: 10.1098/rstb.2006.2000

Morales, M., Mundy, P., and Rojas, J. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behav. Dev.* 21, 373–377. doi: 10.1016/S0163-6383(98)90014-90015

Mundy, P., Block, J., Delgado, C., Pomares, Y., Van Hecke, A. V., and Parlade, M. V. (2007). Individual differences and the development of joint attention in infancy. *Child Dev.* 78, 938–954. doi: 10.1111/j.1467-8624.2007.01042.x

Mundy, P., and Newell, L. (2007). Attention, joint attention, and social cognition. *Curr. Dir. Psychol. Sci.* 16, 269–274. doi: 10.1111/j.1467-8721.2007.00518.x

Oberwelland, E., Schilbach, L., Barisic, I., Krall, S. C., Vogeley, K., Fink, G. R., et al. (2016). Look into my eyes: investigating joint attention using interactive eye-tracking and fMRI in a developmental sample. *Neuroimage* 130, 248–260. doi: 10.1016/j.neuroimage.2016.02.026

Oberwelland, E., Schilbach, L., Barisic, I., Krall, S. C., Vogeley, K., Fink, G. R., et al. (2017). Young adolescents with autism show abnormal joint attention network: a gaze contingent fMRI study. *Neuroimage Clin.* 14, 112–121. doi: 10.1016/j.nicl.2017.01.006

Peirce, J. W. (2008). Generating stimuli for neuroscience using PsychoPy. *Front. Neuroinformatics* 2:10. doi: 10.3389/neuro.11.010.2008

Pfeiffer, U. J., Schilbach, L., Jording, M., Timmermans, B., Bente, G., and Vogeley, K. (2012). Eyes on the mind: investigating the influence of gaze dynamics on the perception of others in real-time social interaction. *Front. Psychol.* 3:537. doi: 10.3389/fpsyg.2012.00537

Pfeiffer, U. J., Schilbach, L., Timmermans, B., Kuzmanovic, B., Georgescu, A. L., Bente, G., et al. (2014). Why we interact: on the functional role of the striatum in the subjective experience of social interaction. *Neuroimage* 101, 124–137. doi: 10.1016/j.neuroimage.2014.06.061

Pfeiffer, U. J., Timmermans, B., Bente, G., Vogeley, K., and Schilbach, L. (2011). A Non-Verbal turing test: differentiating mind from machine in gaze-based social interaction. *PLoS One* 6:e27591. doi: 10.1371/journal.pone.0027591

Pfeiffer, U. J., Timmermans, B., Vogeley, K., Frith, C. D., and Schilbach, L. (2013a). Towards a neuroscience of social interaction. *Front. Hum. Neurosci.* 7:22. doi: 10.3389/fnhum.2013.00022

Pfeiffer, U. J., Vogeley, K., and Schilbach, L. (2013b). From gaze cueing to dual eye-tracking: novel approaches to investigate the neural correlates of gaze in social interaction. *Neurosci. Biobehav. Rev.* 37, 2516–2528. doi: 10.1016/j.neubiorev.2013.07.017

Pinheiro, J. C., and Bates, D. M. (2009). *Mixed-Effects Models in S And S-PLUS*, Reprinted paperback ed. of the 2000 Edn. New York, NY: Springer.

Pönkänen, L. M., Alhoniemi, A., Leppänen, J. M., and Hietanen, J. K. (2011). Does it make a difference if I have an eye contact with you or with your picture? An ERP study. *Soc. Cogn. Affect. Neurosci.* 6, 486–494. doi: 10.1093/scan/nsq068

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Reid, V. M., and Striano, T. (2005). Adult gaze influences infant attention and object processing: implications for cognitive neuroscience. *Eur. J. Neurosci.* 21, 1763–1766. doi: 10.1111/j.1460-9568.2005.03986.x

Reid, V. M., Striano, T., Kaufman, J., and Johnson, M. H. (2004). Eye gaze cueing facilitates neural processing of objects in 4-month-old infants. *Neuroreport* 15, 2553–2555. doi: 10.1097/00001756-200411150-200411125

Risko, E. F., and Kingstone, A. (2011). Eyes wide shut: implied social presence, eye tracking and attention. *Atten. Percept. Psychophys.* 73, 291–296. doi: 10.3758/s13414-010-0042-41

Risko, E. F., Laidlaw, K., Freeth, M., Foulsham, T., and Kingstone, A. (2012). Social attention with real versus reel stimuli: toward an empirical approach to concerns about ecological validity. *Front. Hum. Neurosci.* 6:143. doi: 10.3389/fnhum.2012.00143

Risko, E. F., Richardson, D. C., and Kingstone, A. (2016). Breaking the fourth wall of cognitive science: real-world social attention and the dual function of gaze. *Curr. Dir. Psychol. Sci.* 25, 70–74. doi: 10.1177/0963721415617806

RStudio Team (2016). *RStudio: Integrated Development for R*. Boston, MA: RStudio, Inc.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., et al. (2013). Toward a second-person neuroscience. *Behav. Brain Sci.* 36, 393–414. doi: 10.1017/S0140525X12000660

Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., et al. (2010). Minds made for sharing: initiating joint attention recruits reward-related neurocircuitry. *J. Cogn. Neurosci.* 22, 2702–2715. doi: 10.1162/jocn.2009.21401

Senju, A., and Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Curr. Biol.* 18, 668–671. doi: 10.1016/j.cub.2008.03.059

Senju, A., and Hasegawa, T. (2005). Direct gaze captures visuospatial attention. *Vis. Cogn.* 12, 127–144. doi: 10.1080/13506280444000157

Senju, A., and Johnson, M. H. (2009). The eye contact effect: mechanisms and development. *Trends Cogn. Sci.* 13, 127–134. doi: 10.1016/j.tics.2008.11.009

Tomasello, M., and Carpenter, M. (2005). The emergence of social cognition in three young chimpanzees. *Monogr. Soc. Res. Child Dev.* 70, 107–122. doi: 10.1111/j.1540-5834.2005.00324.x

Tomasello, M., Hare, B., Lehmann, H., and Call, J. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis. *J. Hum. Evol.* 52, 314–320. doi: 10.1016/j.jhevol.2006.10.001

Vernetti, A., Smith, T. J., and Senju, A. (2017). Gaze-contingent reinforcement learning reveals incentive value of social signals in young children and adults. *Proc. R. Soc. B Biol. Sci.* 284:20162747. doi: 10.1098/rspb.2016.2747

Vogeley, K., and Bente, G. (2010). "Artificial humans": psychology and neuroscience perspectives on embodiment and nonverbal communication. *Neural Netw.* 23, 1077–1090. doi: 10.1016/j.neunet.2010.06.003

Wilms, M., Schilbach, L., Pfeiffer, U. J., Bente, G., Fink, G. R., and Vogeley, K. (2010). It's in your eyes–using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience. *Soc. Cogn. Affect. Neurosci.* 5, 98–107. doi: 10.1093/scan/nsq024

Wu, D. W.-L., Bischof, W. F., and Kingstone, A. (2013). Looking while eating: the importance of social context to social attention. *Sci. Rep.* 3:2356. doi: 10.1038/srep02356

Yarbus, A. L. (1967). *Eye Movements and Vision*. Boston, MA: Springer, doi: 10.1007/978-1-4899-5379-7

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 17 December 2019
doi: 10.3389/fnhum.2019.00442

Distinguishing Social From Private Intentions Through the Passive Observation of Gaze Cues

Mathis Jording[1,2]*, Denis Engemann[1,3], Hannah Eckert[2], Gary Bente[4] and Kai Vogeley[1,2]

[1] Cognitive Neuroscience (INM-3), Institute of Neuroscience and Medicine, Research Center Jülich, Jülich, Germany, [2] Department of Psychiatry and Psychotherapy, University Hospital Cologne, Cologne, Germany, [3] Université Paris-Saclay, Inria, CEA, Palaiseau, France, [4] Department of Communication, Michigan State University, East Lansing, MI, United States

Observing others' gaze is most informative during social encounters between humans: We can learn about potentially salient objects in the shared environment, infer others' mental states and detect their communicative intentions. We almost automatically follow the gaze of others in order to check the relevance of the target of the other's attention. This phenomenon called gaze cueing can be conceptualized as a triadic interaction involving a gaze initiator, a gaze follower and a gaze target, i.e., an object or person of interest in the environment. Gaze cueing can occur as "gaze pointing" with a communicative or "social" intention by the initiator, telling the observer that she/he is meant to follow, or as an incidental event, in which the observer follows spontaneously without any intention of the observed person. Here, we investigate which gaze cues let an observer ascribe a social intention to the observed person's gaze and whether and to which degree previous eye contact in combination with an object fixation contributes to this ascription. We varied the orientation of the starting position of gaze toward the observer and the orientation of the end position of a lateral gaze shift. In two experiments participants had to infer from the gaze behavior either mere approach ("the person looked at me") vs. a social ("the person wanted to show me something") or a social vs. a private motivation ("the person was interested in something"). Participants differentially attributed either approach behavior, a social, or a private intention to the agent solely based on the passive observation of the two specific gaze cues of start and end position. While for the attribution of privately motivated behavior, participants relied solely on the end position of the gaze shift, the social interpretation of the observed behavior depended additionally upon initial eye contact. Implications of these results for future social gaze and social cognition research in general are discussed.

Keywords: social gaze, Bayesian multilevel models, ostension, eye contact, communicative intention, gaze cueing

Edited by:
Peter König,
University of Osnabrück, Germany

Reviewed by:
José Luis Ulloa,
University of Talca, Chile
Camila Rosa De Oliveira,
Faculdade Meridional (IMED), Brazil

***Correspondence:**
Mathis Jording
m.jording@fz-juelich.de

Specialty section:
This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

Received: *03 July 2019*
Accepted: *02 December 2019*
Published: *17 December 2019*

Citation:
Jording M, Engemann D, Eckert H, Bente G and Vogeley K (2019) Distinguishing Social From Private Intentions Through the Passive Observation of Gaze Cues. Front. Hum. Neurosci. 13:442. doi: 10.3389/fnhum.2019.00442

INTRODUCTION

The eye region displays emotional and attentional states and is a crucial element in understanding the inner experiences of others (Baron-Cohen et al., 1997; Emery, 2000). This leads to the pivotal role of gaze in social cognition research (Shepherd, 2010) because it informs not only about internal states of persons but also about their relationship to objects or persons in their environment.

Humans process the gaze direction, deduce from it the focus of attention and automatically shift their own attention accordingly. This process is called gaze cueing (Frischen et al., 2007) and is a prerequisite for joint attention, the case in which both persons visually attend the same object. Observing someone looking at objects also informs us about the environment shared by both partners. Accordingly, following someone's gaze changes the perception and processing of jointly attended objects (Becchio et al., 2008); objects, that had previously been looked at by another person are liked more (Bayliss et al., 2006). Gaze following is acquired early in life: 6 month old infants are already able to follow someone's gaze (Senju and Csibra, 2008). Proficiency in gaze following predicts the development of language "theory of mind" capacity (Morales et al., 1998), IQ, self-regulation, social competence and depth of information processing (Mundy and Newell, 2007). It is also believed to be a prerequisite component for reinforcement learning (Vernetti et al., 2017).

A key research question is whether successful gaze processing is an automatic holistic ability, or whether it can be decomposed into distinct cognitive operations, hence, taught and learned. As a clear prerequisite, the gaze angle has to be estimated and the spatial location of the partner's attention has to be inferred from the gaze vector. Compared to great apes and monkeys, humans are especially proficient in this regard (Gibson and Pick, 1963), and the neural implementation of gaze reconstruction has been intensely researched over the past decades (Itier and Batty, 2009).

A second challenge is to discern intentions underlying gaze behavior, which may be explicitly communicative or "social" in the sense that gaze partners want to convey certain information. The "dual function" of gaze comprises the perception of the environment and the signaling of the attentional focus to others (Gobel et al., 2015). I.e., we do not only use the gaze of others as a cue about their attentional focus, but we are at the same time aware that others can deduce our attentional focus from our gaze. Effects of this awareness have been demonstrated impressively in studies showing that participants control their gaze according to its social adequacy when being watched (Risko and Kingstone, 2011). In other words, humans are forced to actively avoid undesired communication by controlling their eye gaze in social contexts. Likewise, when observing another person, this person's gaze might be driven by self-centered interests or it might be an attempt to communicate or to express a "social" intention. Thus when deducing the other's intentions, perceivers have to distinguish between "private" and "communicative" intentions (Walter et al., 2004). It can be expected that this distinction fundamentally affects our relationship toward the other person. Walter et al. (2004) could show that, during mentalizing, the processing of private and communicative intentions rely on distinct neural mechanisms, even if the communicative actions are not directed toward the observer.

Csibra and Gergely (2009, 2011) speculate that humans use eye contact as an "ostensive" signal to announce situations in which they want to show or teach something to others. Being gazed at by another person is a powerful social cue to which most humans are highly sensitive (von Griinau and Anston, 1995; Senju and Johnson, 2009), and eye contact is supposed to signal communicative intents (Kleinke, 1986). Conversely, according to Csibra and Gergely (2009, 2011), infants have an innate sensibility to ostensive cues which allows them to generalize their experience in these situations in order to fully benefit from their teacher. Preceding communication indeed has been shown to facilitate subsequent gaze cueing and gaze following already in 4–6 month old infants (Farroni et al., 2003). This mechanism might also explain the strong ontogenetic link not only between gaze following and joint attention, but also between the mental and cognitive development.

Here, we present two studies that explore the link between gaze direction processing and communicative or "social" affordances. We investigate the principles of how humans deduce the attentional focus from others' gaze with regard to the tension between private and communicative or "social" intentions. The motivation for Study 1 was to study the role of eye contact in reducing the ambiguity of gaze and to identify the parameters that allow to interpret the gaze behavior of others as ostensive, i.e., a special case of communicative intention that bridges the gap between person and environment. Specifically, we aimed at the difference between situations in which we experience an interacting partner as being interested in us by visually attending to us in contrast to situations in which the partner is actively trying to communicate with us about something in the outside world by a rudimentary form of joint attention. The observation of distinct patterns of observed gaze in the two conditions lead us to the question in Study 2, whether and how participants distinguish aforementioned communicative intentions from situations in which the partner is experienced as being "privately" interested in something without involving and addressing the perceiver.

As the basic design of both studies, participants watched short videos of a virtual character (VC) looking at the participant with different degrees of vertical deviations, ranging from direct gaze (i.e., eye contact) to different degrees of downward averted gaze, before shifting the gaze to the left or to the right with different degrees of lateral deviations. (For simplicity, we will refer to the starting position of initial gaze as "initial position" and to the gaze shift to the left or to the right as "shift amplitude"). Subsequently, participants had to report their experiences based on explicit statements (see **Figure 1**). We used VCs as stimulus material, as they combine high experimental control with ecological validity (Vogeley and Bente, 2010) and are well suited for the investigation of non-verbal communication (Pfeiffer et al., 2013, 2014; Georgescu et al., 2014; Jording et al., 2019).

In the first study, we investigated the difference between situations in which participants had the impression of been looked at by the VC ("LOOK" condition) and situations in which they had the impression that the VC was trying to show them something ("COM," e.g., "communicative," condition). Besides the aforementioned empirical question, a second goal of this first study was to ensure the validity of our stimuli and the overall methods. The sensitivity of human observers to the visual stimulus of eyes directed at them is already well established (Senju and Johnson, 2009) and VCs were shown to reliably

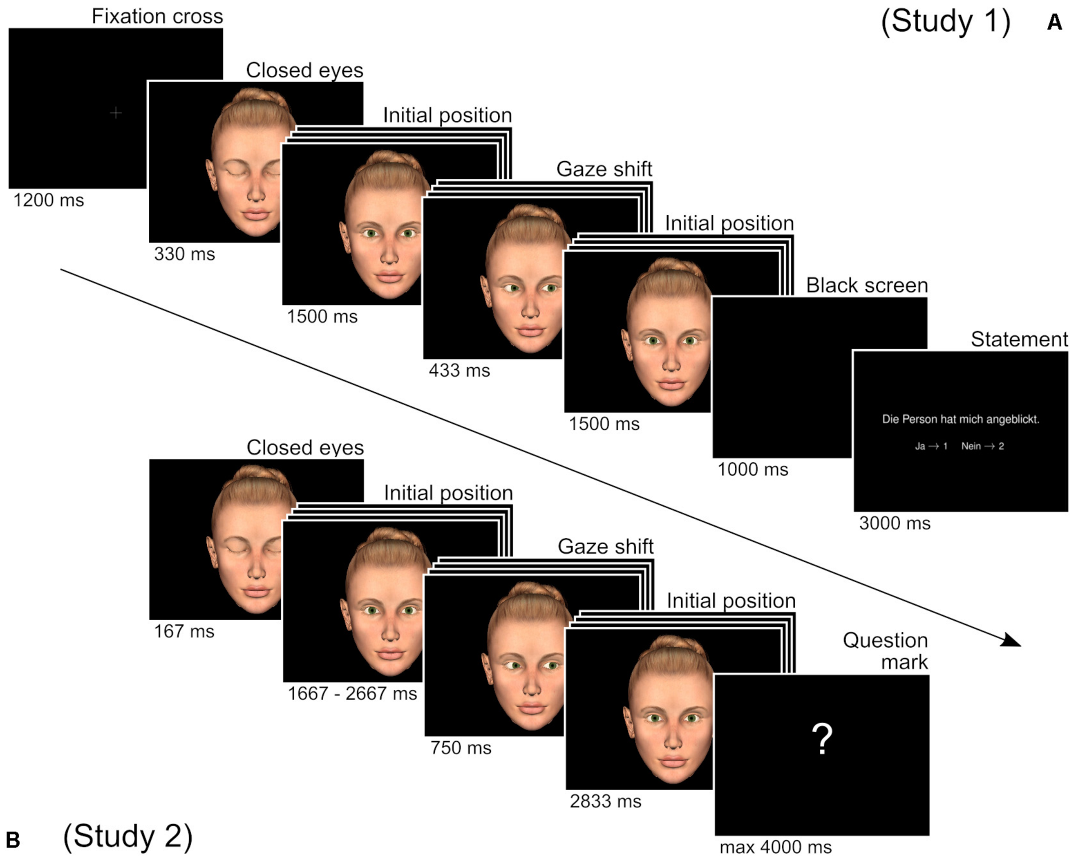


FIGURE 1 | Course of one trial of Study 1 (A) and Study 2 (B). The stack of images for initial position and shift position indicate that in each trial, one out of four possible images was displayed. Note that after the shift, the VC always returned to the same initial position it had started from. In this example, the initial position 1 (direct gaze) and the shift position 4 are depicted. The question mark indicated the prompt for participants to give their ratings.

induce the impression of social presence (Bente et al., 2007). Our stimuli can elicit the feeling of being in the attentional focus of or being addressed by the VC. Therefore, results of this first study serve as a test of these properties of the stimuli. This first study was conducted as an internet-based online survey to maximize sample size and account for possible variability in the general population. In the second study with a new sample of participants, we again studied COM in comparison to the situation in which the VC was merely privately interested in something without any social intention ("PRIV" condition). This second study had a repeated measures design and was conducted in a laboratory setting, increasing experimental control of environment and participant specific factors.

We expected the impression of being looked at to be dependent solely on the degree to which the initial gaze is directed toward the participants but not on subsequent outward-directed behavior. In COM, available evidence in the field suggested an influence of preceding eye contact for the impression of communicative intentions as well. Considering that participants were asked whether the other wanted to show them something located in the outside world, we also expected an influence of the subsequent gaze shift during COM. However, this situation by definition requires a triadic interaction between two interactants and another object in the environment. Therefore, we expected high agreement rates only for situations with direct gaze and large shift amplitudes. During PRIV, we expected an influence of the shift amplitude only. However, it was also interesting to see whether preceding eye contact might have an adverse effect. Should participants understand private and communicative intentions as mutually exclusive, they should take eye contact as an indicator of the latter, leading to an impediment of the impression of mere personal interest.

MATERIALS AND METHODS

Study 1

Participants for Study 1

Out of 555 participants, 403 participants completed the online survey. In 11 cases videos were not presented correctly, resulting in 392 remaining participants (257 female; age ranging from 17–70 years, $M = 30$, SD = 10.63). Participants were recruited via mailing lists from different German universities (University of Cologne, University of Münster, University of Bayreuth) and gave their informed consent prior to participation. There were no further exclusion criteria.

Stimuli for Study 1

One female and one male VC were created with Poser for Apple Mac OS X (Poser 8, Smith Micro Software, Inc., Columbia, SA, United States). For both VCs images were created for four different initial gaze positions and for four different gaze shift targets in two different directions. Initial gaze positions were equidistantly positioned on a central vertical line, ranging from direct gaze to clearly averted gaze. Positions after gaze shifts were equidistantly located on a horizontal line slightly below the eye level, ranging from slight central deviation up to the maximal still realistic and lifelike appearing deviation, both for the right and the left side. From these images we approximated the deviation of the visual angle from direct gaze (initial position 1) by measuring for all images the position of the iris in relation to its position in the direct gaze image. On this basis we computed angles, taking 22 mm as the average diameter of the human eye (Bekerman et al., 2014) and 12 mm as average diameter of the human iris (Thainimit et al., 2013). Averaged between VCs, the initial positions vertically deviated approximately equidistantly from direct gaze by 0°, 3°, 8°, and 12°. VC-averaged gaze positions after the shift lay on a plane 6° vertically below the eye level, horizontally deviating from direct gaze approximately equidistantly by 5°, 9°, 14°, and 18°. (For examples of all initial positions and gaze shift images and the exact values of the degree of aversion, please refer to the **Supplementary Material**.) Images of initial positions and gaze shifts were then combined to flash videos by the python 2.6 based video tool "ffmpeg 0.7.8." For both sexes of VCs videos were created for each combination of four different initial positions and four different shift amplitudes to both sides, resulting in 16 videos of gaze shifts to the right and 16 videos for gaze shifts to the left per VC and a total of 64 videos. Each video started with showing a fixation cross for 1200 ms. Afterward the VC appeared, having his/her eyes closed for 330 ms before he/she subsequently opened the eyes and looked toward the initial position for 1500 ms, then shifted toward the target for 433 ms, before returning to the initial position for 2000 ms. Afterward the screen went black for 1000 ms, before the statement and response buttons were displayed for 3000 ms as a reminder at the end of the video (see **Figure 1A** for an illustration and **Supplementary Videos S1–S4** for examples of the trial course).

Task for Study 1

Each participant watched videos of either the female or the male VC for all 16 different combinations of gaze initial positions and shift amplitudes to the left or to the right in randomized order exactly once. After each video participants had to rate the VCs behavior according to statements randomly assigned in the beginning of the experiment. Statements were either "the person looked at me" (German original: "Die Person hat mich angeblickt") or "the person wanted to show me something" (German: "Die Person wollte mir etwas zeigen"), to which participant had to respond per button press in a binary choice ("yes" or "no").

Setup and Design for Study 1

The survey was presented via the online survey tool Unipark (Questback GmbH, Cologne, Germany). During the survey, participants were informed about the procedure, the voluntary nature of their participation and the opportunity to withdraw from the study at any point in time and without providing any reasons for their decision. They further had to state their age and sex before they were pseudo-randomly assigned to one of the two VCs and one of the two rating statements. After that, participants were told which statement they had to answer and whether they would see a female or a male character. Participants were then presented with videos for all 16 combinations of initial positions and shift amplitudes in a pseudorandomized order with shifts randomly either to the right or the left. After each video the screen turned black before the statement was presented together with the binary response options (button "1" for "Yes" and button "2" for "No"). The next trial then started after the participants had given their answers.

Statistics for Study 1

The effect of different gaze shifts (initial position and shift amplitude) on the ascription of different intentions to the VC (conditions) were analyzed in a multilevel model with an inverse logit link function, in which we considered individual differences between the participants' average responses through varying intercept coefficients. Importantly, we considered the statement as experimental condition and hence constructed a joint model for both statements instead of two separate models. The model focuses on the interaction between the statement and eye gaze behavior. This approach has enabled explicitly modeling statement-specific-biases, e.g., due to difficulty or individual preferences, while, at the same time subjecting the estimated differences between the effects to statistical control through shrinkage priors (see below). The resulting logistic regression model can be expressed as:

$$y_i \sim Binomial(n = 1, p = \hat{y}_i)$$

$$\hat{y}_i = logit^{-1}(\alpha_j[i] + T[i] * \beta)$$

Where α_j is the individual intercept for each subject, T is a matrix of treatment effects, and β the unknown parameter vector that has to be learnt from the data. The treatment effects are the statement, the vertical initial gaze position and the horizontal amplitude of the gaze shift, covering all main effects as well as second and third order interactions. The statement was dummy-coded with a 0–1 predictor. We included the eye gaze as continuous predictor after z-scoring. No prior information concerning effect sizes of the initial gaze position or shift amplitude were available. We hence used the non-informative default priors from the "brms" package (Bürkner, 2017, 2018) according to which coefficient are centered around zero. These priors are shrinkage priors and are conservative. Shrinkage is used in statistics to improve generalization to new data can be thought of correcting initial estimates by pushing them toward zero. The amount of shrinkage fades out as the sample size increases. For the prior for the population variance component σ_j of the individual intercepts, we kept the conservative default prior that puts most probability mass on smaller values close to zero.

$$\beta \sim \mathit{student's}\, t(df = 3,\ center = 0,\ \sigma^2 = 10)$$

$$\alpha_j \sim \mathit{student's}\, t(df = 3,\ center = 0,\ \sigma^2 = \ \sigma_j)$$

$$\sigma_j \sim \mathit{half\text{-}student's}\, t(df = 3,\ center = 0,\ \sigma^2 = \ \sigma_j)$$

Note that the population variance parameter σ_j uses the upper half of the student-t distribution due to the constraint that the variance cannot be negative. Also note that σ_j is a hyper-parameter and has to be estimated from the data. Here, it controls how much the model trusts the individual intercept estimates σ_j and to which extent these will be corrected by shrinkage toward the global intercept. Smaller values for σ_j would produce stronger shrinkage. This is a core feature of the multilevel model and is also referred to as partial pooling (Gelman, 2006).

We performed prior predictive checks to ensure that the priors are approximately uninformative on the scale of the model predictions after the inverse logistic link function. Analysis revealed that the results were insensitive to the choice of the prior due to the size of the data set. Data were analyzed using the "rstan" (Stan Development Team, 2018) and "brms" (Bürkner, 2017, 2018) packages for the programing language R for statistical computing (R Development Core Team, 2008) and RStudio (R Studio Team, 2016). Model fitting was performed using a Hamilton Markov chain Monte Carlo algorithm (Hoffman and Gelman, 2014). Models were run with 1000 warmup samples and 1000 iterations in total, using four chains, yielding 4000 draws from the approximated posterior distribution. Successful convergence was assessed based on the potential scale reduction factor $\hat{R}$, also known as the Gelman-Rubin statistic. $\hat{R}$, was found to be acceptably close to 1.0 (±0.1) for every model (see **Supplementary Table S1**). Posterior distributions were visually compared to observed data in order to check consistency.

Study 2

Participants for Study 2

34 subjects (19 female; age range 21–54, M = 28.88, SD = 5.82; not out of the sample from Study 1) participated in this experiment. None of these participants met any of the exclusion criteria (depressive symptoms as indicated by BDI scores: M = 3.79, range = 0–17, cut-off ≥ 19; autistic traits as indicated by AQ scores: M = 10.42, range = 2–19, cut-off ≥ 32; general cognitive impairments as indicated by MWT: M = 112.59, range = 97–136, cut-off < 70, or KAI, M = 124,24, range = 100–143, cut-off < 70) so that all participants were included for further analysis. The mean empathy score of the resulting sample as indicated by the SPF was M = 40.64, range = 30–49. Participants were recruited via mailing lists from the University of Cologne and gave their informed consent before participating.

Stimuli for Study 2

The same VC pictures were used as in Study 1. Instead of beforehand creating animated videos, as in Study 1, images were now combined to animations within the presentation software (Python 2.6), allowing for jittering of presentation durations. As in Study 1, animations of both VCs could be presented displaying gaze shifts for all 16 possible combinations of initial positions and shift amplitudes to both directions (left and right), resulting in a total of 32 different gaze shifts per VC. Each video sequence started with the VC having its eyes closed for 167 ms (10 frames) before opening them and looking toward the gaze initial position for 1667–2667 ms (100–160 frames). Afterward the VC gaze shifted and then stayed at the new location for 750 ms (45 frames) before returning to the initial location at the end of the video for another 2833 ms (170 frames). Subsequently, a screen showing a white question mark in front of a black background requested the participants to give their answer for a maximum of 4000 ms. (Please refer to **Figure 1B** for an illustration and **Supplementary Videos 5–8** for examples of the trial course).

Task for Study 2

In accordance with Study 1, participants, after having watched a gaze shift performed by the VC, had to rate the VCs behavior according to one of two different statements per trial. The statements were either "the person wanted to show me something" (German: "Die Person wollte mir etwas zeigen") or "the person was interested in something" (German original: "Die Person interessierte sich für etwas"). Again, participants had to respond per button press in a binary choice ("Yes" or "No"), for which they had 4 s before the next trial would start.

Setup and Design for Study 2

Before the experiment started, participants general cognitive level was assessed by two tests: KAI (Lehrl et al., 1991) and MWT-B (Lehrl, 2005). The experiment was conducted on a Lenovo ThinkPad T410 (Intel Core i5-520 M, 2,4 Ghz, 4GB RAM; OS: Ubuntu Linux 12.4 LTS) and displayed on a Tobii T60 Eye Tracker (60 Hz refresh rate, 1280 × 1024 px resolution) with responses given via keypad buttons and instructions presented on the screen. For the experiment, two blocks of trials (one block per statement) were presented in a pseudorandomized fashion. In each block, the participant watched all 64 gaze shifts (four initial positions × four shift amplitudes × two directions × two VCs) resulting in a total of 128 trials per participant over the whole experiment and a total duration of approximately 20 min. Before the experiment, KAI (Lehrl et al., 1991) and MWT-B (Lehrl, 2005) were conducted to rule out general cognitive impairments. After the experiment participants completed BDI (Beck et al., 2001), and AQ (Baron-Cohen et al., 2001) to rule out depressive and autism-like syndromes, respectively. In addition, participants filled out the empathy inventory SPF (Paulus, 2009) to potentially allow the matching with patient samples in future clinical studies.

Statistics for Study 2

The same statistical procedures where applied as in Study 1 (for $\hat{R}$ values see **Supplementary Table S2**). Note that the multilevel approach has allowed us to use the same model specification for Study 2, as this kind of model is robust to the structure of repeated observations and can be applied to a wide array of between or within-subject designs (see McElreath, 2016, Chapter 12, box on pp. 371 for discussion).

RESULTS

Interpreting multilevel models solely based on their coefficients is known to be notoriously difficult, especially for generalized linear models with non-Gaussian probability models (Ai and Norton, 2003). As is common practice, we therefore considered posterior predictions (**Figure 2** for Study 1; **Figure 4** for Study 2) in addition to model coefficients (**Figure 3**, Study 1; **Figure 5**, Study 2). The posterior predictions contain the uncertainty of the model and can be readily interpreted in terms of the probability of the responses given the model and the data. They conveniently support statistical inference and can be analyzed in terms of percentiles or subtracted from another to form contrasts. For the effect of the individual

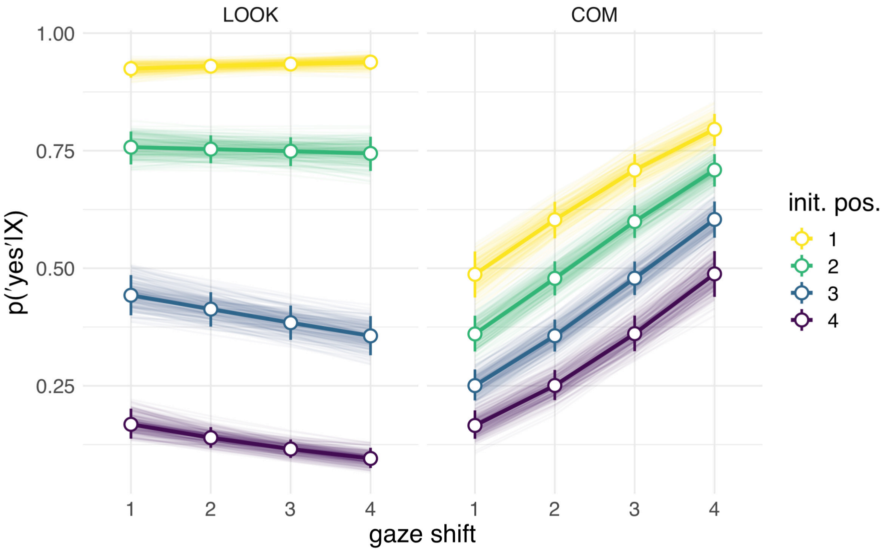


FIGURE 2 | Posterior predictions of the influence of initial position („init. pos.“) and gaze shift amplitude („gaze shift“) in Study 1 in the LOOK condition (“the person looked at me”) and the COM condition (“the person wanted to show me something”). For the initial position, „1“ corresponds to direct gaze and „4“ to a maximally (vertically) averted position. For the shift amplitude, „1“ corresponds to the smallest and „4“ to the largest possible shifts.

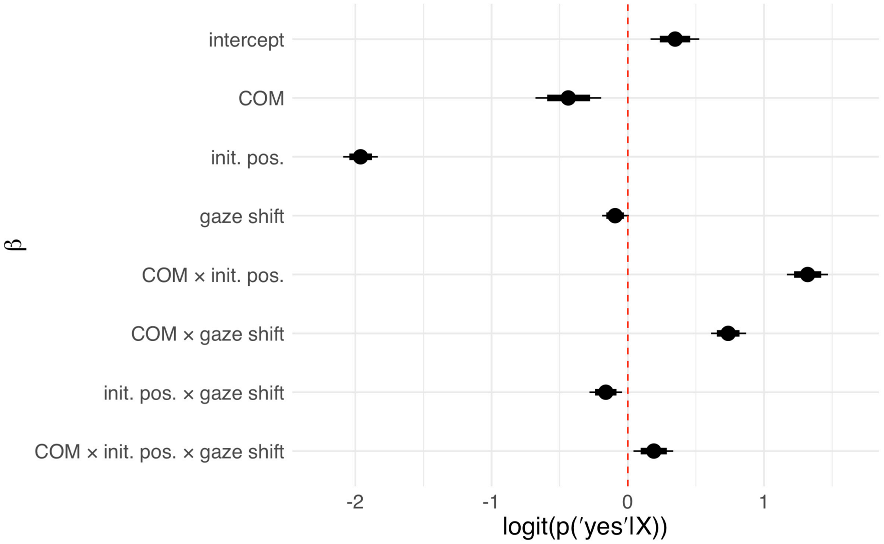


FIGURE 3 | Coefficients sampled from the approximate posterior distribution in Study 1 for the influence of condition, initial position, shift amplitude, and their respective interactions. Circles depict the posterior mean, horizontal bars and lines denote the 80 and 95% posterior compatibility intervals, respectively. The COM coefficient describes the effect of the COM condition in contrast to the LOOK condition. The coefficient for initial positions depicts the stepwise effect of increasing aversion from direct gaze in the initial position (farther from direct gaze). The coefficient of shift amplitude depicts the stepwise effect of increasing the shift amplitude. For additional statistics see **Supplementary Table S1**; Note that although not apparent here, the 95% confidence interval of the gaze shift coefficient does include zero.

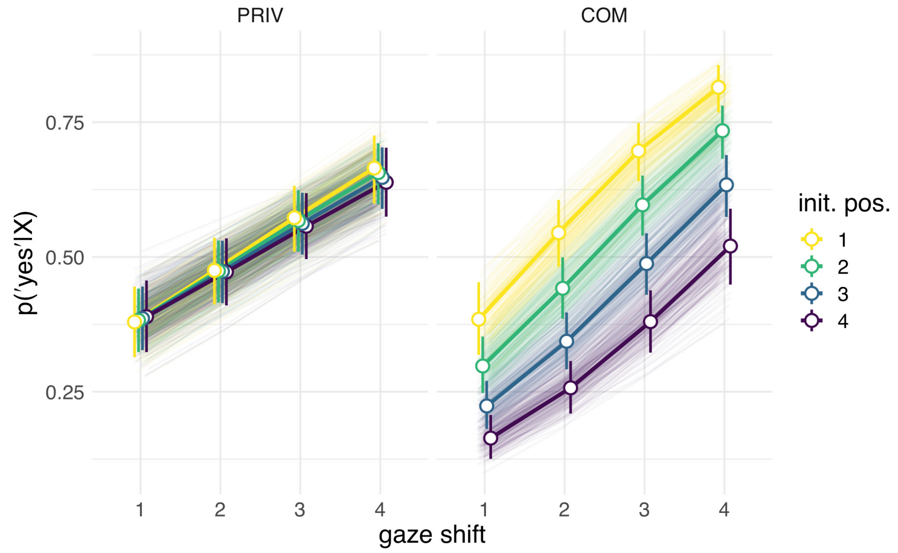


FIGURE 4 | Posterior predictions of the influence of initial position („init. pos.") and gaze shift amplitude („gaze shift") in Study 2 in the PRIV condition ("the person was interested in something") and the COM condition ("the person wanted to show me something"). For the initial position, „1" corresponds to direct gaze and „4" to a maximally (vertically) averted position. For the shift amplitude, „1" corresponds to the smallest and „4" to the largest possible shifts.

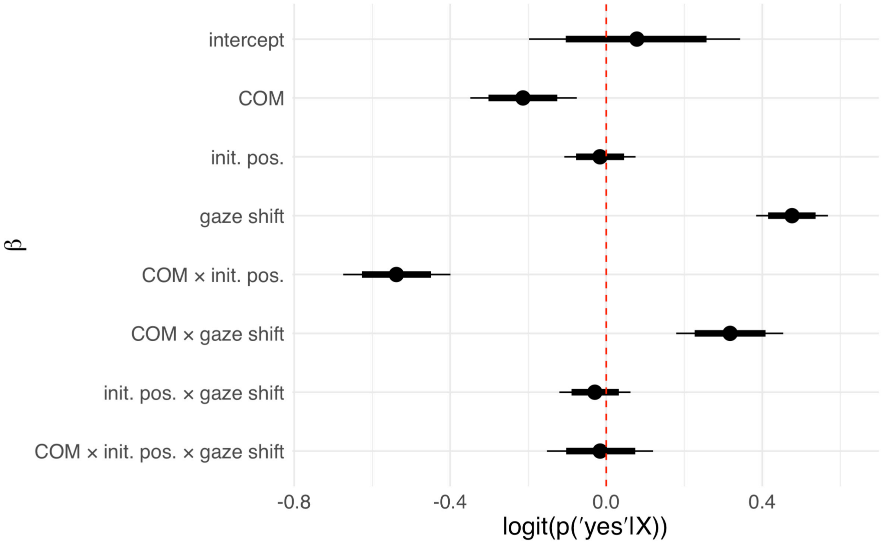


FIGURE 5 | Coefficients sampled from the approximate posterior distribution in Study 2 for the influence of condition, initial position, shift amplitude, and their respective interactions. Circles depict the posterior mean, horizontal bars and lines denote the 80 and 95% posterior compatibility intervals, respectively. The COM coefficient describes the effect of the COM condition in contrast to the PRIV condition. The coefficient for initial positions depicts the stepwise effect of increasing aversion from direct gaze in the initial position (farther from direct gaze). The coefficient of shift amplitude depicts the stepwise effect of increasing the shift amplitude. For additional statistics see **Supplementary Table S2**.

predictors, beta coefficients (as well as the respective 80 and 95% posterior probability distribution intervals) are reported in the **Figures 3**, **5**, additional statistics can be found in the **Supplementary Tables S1**, **S2**. This approach was chosen in order to increase the comparability to traditional reports of frequentist statistical methods with 0.05 significance levels. The intercepts for Study 1 and 2 refer to the COM condition, coefficients for the LOOK condition (Study 1) or the PRIV condition (Study 2) describe the change in coefficients compared to this intercept.

Study 1

In Study 1 (online study) 198 participants (134 female, 64 male; age: 17–66 years, *M* = 29.37, SD = 10.69) participated in the LOOK condition and 194 participants (123 female, 71 male; age: 18–70 years, *M* = 30.07, SD = 10.59) participated in the COM condition. We compared posterior predictions for agreements to LOOK ("the person looked at me") and COM ("the person tried to show me something") statements (**Figure 2**). Posterior predictions revealed that participants discriminated the two conditions based on the two gaze dimensions, initial position and shift amplitude (**Figure 2**). In the COM condition, the effect of initial position as well as shift amplitude had substantial effects with the probability of agreement to the statement "the person tried to show me something" increasing with initial positions closer to eye contact and larger shift amplitudes. In comparison, in the LOOK condition, the effect of the initial position was even more pronounced while the shift amplitude did not show any considerable effect on the probability of "the person looked at me" statement. In addition, a slight tendency to higher overall agreements to the LOOK compared to the COM statements is visible. These results are reflected in the configuration of the model coefficients (**Figure 3**) which uncovered higher order interaction effects between condition and the dimension of gaze shifts.

Study 2

In Study 2 (Lab Study) all 34 subjects participated in both conditions (COM and PRIV) in a repeated measures design. Here, we tested whether results from the COM condition in Study 1 could be replicated and how they would compare to the PRIV condition. In posterior predictions (**Figure 4**) for the COM condition the same pattern as in Study 1 arose with the probability of agreeing with the statement "the person tried to show me something" increasing with initial positions closer to eye contact and with larger shift amplitudes. Corroborating results of Study 1, no considerable interaction effect between initial position and shift amplitude was observed. In comparison, posterior predictions for the PRIV condition revealed that the overall tendency to agree with the statement "the person was interested in something" was slightly higher. Larger shift amplitude enhanced the probability of agreement even further, although this effect was less pronounced in PRIV compared to COM. Neither the initial position nor the interaction between initial position and shift amplitude had considerable effects in PRIV. Results correspond to the configuration of model coefficients (**Figure 5**), which uncovered simple but no higher order interactions.

DISCUSSION

The present study focused on the interplay of person-related and environment-related aspects of gaze behavior and how they influence our tendency to ascribe communicative or "social" and "private" intentions. The impression of being looked at (LOOK) has proved to be highly relying on initial eye contact for only in the conditions of direct gaze (or only slightly diverted gaze) ratings reached at least 75% agreement rates, while in cases of more diversion, agreement decreased substantially. Given the high sensitivity of humans to eye contact (von Griinau and Anston, 1995; Senju and Johnson, 2009) and its close link to intimacy (Argyle and Dean, 1965) this finding appears highly plausible. The amplitude of the subsequent gaze shift had no decisive influence, which corresponded also with our expectation.

The communicative condition (COM) revealed substantially the same results in the online study as in the laboratory study. Here, direct gaze or starting points close to it during the initial gaze and large gaze shifts significantly fostered the impression of being shown something. This matches the role of eye contact conveying communicative intentions (Kleinke, 1986) and nicely fits accounts of eye contact being used as ostensive cue. However, the ostensive situation also extends beyond the dyadic interaction of the two persons to the outside world. This is represented in the increasing effect of the assumed goal-directedness of the gaze shift. In other words, gaze contact with the viewer is only one component, the other component that makes this gaze behavior ostensive, is obviously the gaze shift directed toward an invisibly target in the environment. This result also ties in with other findings showing that infants as young as 9 month are not only sensitive to ostensive gaze cues, but they also expect object directed gaze shifts in these situations (Senju et al., 2008). Similarly, we had expected that participants would experience communicative intentions only when the triadic nature of the situation was apparent in the agents' gaze behavior. Accordingly, we expected to find an interaction effect between the degree of eye contact and shift amplitude for the COM condition. However, this interaction effect proved to be negligible compared to the observable main effects. Thus, in our initial hypothesis we overestimated the component to which participants considered contextual factors when inferring communicative intentions. The question therefore remains, to which extent the effect of ostensive signals facilitating gaze cueing can be ascribed to more fundamental levels of processing. When investigating the reallocation of attention in a similar situation, Bristow et al. (2007) were able to identify a corresponding interaction effect. BOLD-responses in the parieto-frontal attentional network indicated a stronger reallocation of attention for the observation of gaze shifts toward empty space vs. an object when the observed face had previously looked at the participant in contrast to an averted gaze condition. The authors assumed that the enhanced (visual) saliency of eyes directed at the viewer might have increased the gaze cueing effect.

When participants had to rate whether or not the VC appeared to be interested in something (PRIV), only shift amplitude had a notable effect with larger gaze shifts eliciting higher approval rates. We assume that participants tended to perceive small gaze shifts as still directed toward them. Despite the human general acuteness in retracing gaze vectors and directions, they show a surprising tolerance when identifying gaze directed at them with deviations up to several degrees (Gibson and Pick, 1963; Jenkins et al., 2006; Mareschal et al., 2013). Interestingly, this tendency is even stronger for participants that had experienced social exclusion prior to the experiment (Lyyra et al., 2017). We, however, did neither induce or ask explicitly for the experience of social exclusion.

It makes sense that participants, when asked whether the other one was interested in something, assumed this something in the outside world and took more decisive gaze shifts as reflecting this interest. In general, humans, when observing another persons' gaze, express some flexibility not only with regard to gaze directed at them, but also when it is directed at objects. We perceive a person as looking directly toward an object even in case of an actual divergence between gaze vector and object (Lobmaier et al., 2006). Unfortunately, research on the effect of the target position and shift amplitude in gaze cueing is still sparse. To the best of our knowledge, only one study investigated the gaze cueing effect as a function of the cued position, reporting higher effects for more distant positions (Qian et al., 2013). Our data now suggest, that when gaze shifts were more pronounced, participants more strongly imagined the existence of objects in their shared environment, even though not visible to them. However, due to the still insufficient knowledge about the underlying mechanisms this notion remains speculative.

It is interesting that the initial gaze does not influence the judgment. Even when initially eye contact was established, this did not impede the impression of privately motivated behavior so that the interpretation of the same behavior either as communicative or as private crucially depends on the instruction or the "mindset." Obviously, private and communicative intentions are not mutually exclusive, a person can be interested in something and therefore try to show it to others. However, at least in this highly reductionistic quasi-"social" context, participants did not or were not able to distinguish between those two situations.

Taken together, results corroborate that the combination of mere eye contact and lateral gaze shift together can already signal communicative intentions in a very robust way and can serve as powerful ostensive cue. However, data suggest that eye contact itself and even in combination with the subsequent gaze shift are not sufficient to biuniquely discern intentions from social gaze. The impression of communicative intentions was most prevalent in, but not limited to, the most profiled triadic situations, defined by initial eye contact and large gaze shift amplitudes. This is in line with results showing that ostensive gaze cues do not necessarily seem to be a prerequisite for gaze following in infants (Szufnarowska et al., 2014; Gredebäck et al., 2018). Conversely, eye contact did not inhibit the impression of private intentions. With regard to the differentiation between communicative and private intentions, this means that eye contact neither seems to constitute a highly predictive nor selective signal. Thus, the question remains, which other signals or processes might be used discern intentions from gaze.

Here, the highly reductionist approach of this study clearly reaches its limits. While it was warranted for elucidating the relationship between the most basic aspects of ostensive gaze behavior, its limitations have to be considered as well. First: Non-verbal communication in general was already pointed out to have a high procedural and dimensional complexity meaning that individual non-verbal cues are not isolated units but always part of a stream of cues from different non-verbal channels (Vogeley and Bente, 2010). Regarding the investigation of gaze behavior it is thus advisable not to limit the analysis to short chunks of gaze communication and potentially to include other non-verbal channels as well (Jording et al., 2018). Second: The context or environment has to be taken into account when investigating gaze processing (Hamilton, 2016). Adding and systematically varying objects to the setup as a focus point for the ostensive gaze cues would thus constitute another interesting variation of this study. Third: Closely linked to environmental aspects are factors regarding our knowledge about the other person. Although gaze cueing and gaze following can happen automatically, it is also influenced by our perception and beliefs about the other person as well as our relationship toward this person (Gobel et al., 2017). Thus, systematically manipulating the participants believes about of the observed agent (e.g., personality or preferences) might influence their interpretation of the observed gaze behavior.

CONCLUSION

In conclusion, although the two studies on gaze behavior presented here are highly minimalistic, they nevertheless substantially deepen our understanding of the powerful potential of social gaze in initiating interactions, referencing and displaying attention and thus allow a glimpse through the "window into social cognition" that social gaze can provide (Shepherd, 2010). Eye contact has again been proven to be a powerful tool in imparting communicative intents and fostering the impression that someone else is actively trying to show us something. However, it also becomes evident that eye contact itself is obviously not sufficient to discern intentions from social gaze biuniquely. Humans most likely make use of additional, e.g., temporal characteristics of gaze or they take other non-verbal or verbal signals into account; further investigations on this topic are therefore warranted. In practice, this study can inform us about the fundamental processes that underlie the perception and potentially production of gaze behavior and their functional roles in communication. Technically, these insights may help develop applications in the field of interaction and communication sciences by making use of anthropomorphic virtual agents and humanoids (Pfeiffer et al., 2013). In order for cognitive robots to become accepted as interaction partners by humans they have to share the human ability to generate and interpret informative gaze behavior as a two-way communicative act (Pönkänen et al., 2011; Gobel et al., 2015; Jording et al., 2018). A more thorough understanding of how humans convey and ascribe intentions as supplied here is therefore essential. In the long-run this approach might then also foster the development of more sophisticated agent-based diagnostic and therapeutic instruments for communication disorders like autism spectrum disorders (Georgescu et al., 2014).

DATA AVAILABILITY STATEMENT

Data will be made available in a public repository upon publication and till then can be accessed via https://osf.io/avu5w/?view_only=bd507466e05544589eac294c33253e8c.

ETHICS STATEMENT

This study followed the WMA Declaration of Helsinki (Ethical Principles for Medical Research Involving Human Subjects) and was presented to and approved by the Ethics Committee of the Medical Faculty of the University Hospital Cologne, Germany.

AUTHOR CONTRIBUTIONS

All authors substantially contributed to the conception of the work. DE programed the code for data collection in the online and in the lab study. HE recruited participants and conducted the lab experiment. MJ and DE conducted the statistical analyses. MJ drafted the manuscript. HE, DE, GB, and KV revised the manuscript critically.

FUNDING

KV received funding from the Federal German Ministry of Education and Research (Grant No.: 16SV7244) and from the EC, Horizon 2020 Framework Program, FET Proactive (Grant Agreement ID: 824128).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2019.00442/full#supplementary-material

FIGURE S1 | Illustration of the female avatar's eyes and the measurement of the iris' position for gaze angle calculation. Middle column: Eye section from the female avatar stimuli for the initial gaze position (top four) and the position after the gaze shift (bottom four). Red circles with a centered cross mark the position of the iris as measured for the calculation of the gaze angle. Right column: lateral and horizontal deviations of the gaze angle from direct gaze. Note that depicted here are only gaze shifts to the left side; for shifts to the right side avatar stimuli were mirrored.

FIGURE S2 | Illustration of the male avatar's eyes and the measurement of the iris' position for gaze angle calculation. Middle column: Eye section from the female avatar stimuli for the initial gaze position (top four) and the position after the gaze shift (bottom four). Red circles with a centered cross mark the position of the iris as measured for the calculation of the gaze angle. Right column: lateral and horizontal deviations of the gaze angle from direct gaze. Note that depicted here are only gaze shifts to the left side; for shifts to the right side avatar stimuli were mirrored.

TABLE S1 | Coefficients sampled from the approximate posterior distribution in Study 1 for the influence of the COM condition, initial position, shift amplitude, and their respective interactions. The COM coefficient describes the effect of the COM condition in contrast to the LOOK condition; init. pos. depicts the stepwise effect of increasing aversion from direct gaze in the initial position (farther from direct gaze); gaze shift depicts the stepwise effect of increasing the shift amplitude. Reported are estimates (Estimate) and estimated errors (Est.Error) for the coefficients, the lower (l-95% CI) and the upper (u-95% CI) border of the 95% posterior compatibility intervals, the effective sample size (Eff.Sample) and the potential scale reduction factor $\hat{R}$ or Gelman-Rubin statistic ($\hat{R}$).

TABLE S2 | Coefficients sampled from the approximate posterior distribution in study 2 for the influence of the COM condition, initial position, shift amplitude, and their respective interactions. The COM coefficient describes the effect of the COM condition in contrast to the PRIV condition; init. pos. depicts the stepwise effect of increasing aversion from direct gaze in the initial position (farther from direct gaze); gaze shift depicts the stepwise effect of increasing the shift amplitude. Reported are estimates (Estimate) and estimated errors (Est.Error) for the coefficients, the lower (l-95% CI) and the upper (u-95% CI) border of the 95% posterior compatibility intervals, the effective sample size (Eff.Sample) and the potential scale reduction factor $\hat{R}$ or Gelman-Rubin statistic ($\hat{R}$).

VIDEO S1 | Example Study 1_init.1_shift.4.

VIDEO S2 | Example Study 1_init.2_shift.3.

VIDEO S3 | Example Study 1_init.3_shift.2.

VIDEO S4 | Example Study 1_init.4_shift.1.

VIDEO S5 | Example Study 2_init.1_shift.4.

VIDEO S6 | Example Study 2_init.2_shift.3.

VIDEO S7 | Example Study 2_init.3_shift.2.

VIDEO S8 | Example Study 2_init.4_shift.1.

REFERENCES

Ai, C., and Norton, E. C. (2003). Interaction terms in logit and probit models. *Econ. Lett.* 80, 123–129. doi: 10.1016/S0165-1765(03)00032-36

Argyle, M., and Dean, J. (1965). Eye-Contact, distance and affiliation. *Sociometry* 28, 289–304. doi: 10.2307/2786027

Baron-Cohen, S., Wheelwright, S., and Jolliffe, T. (1997). Is there a "Language of the Eyes"? Evidence from normal adults, and adults with autism or asperger syndrome. *Vis. Cogn.* 4, 311–331. doi: 10.1080/713756761

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., and Clubley, E. (2001). The autism-spectrum quotient (AQ): evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. *J. Autism Dev. Disord.* 31, 5–17.

Bayliss, A. P., Paul, M. A., Cannon, P. R., and Tipper, S. P. (2006). Gaze cuing and affective judgments of objects: i like what you look at. *Psychon. Bull. Rev.* 13, 1061–1066. doi: 10.3758/BF03213926

Becchio, C., Bertone, C., and Castiello, U. (2008). How the gaze of others influences object processing. *Trends Cogn. Sci.* 12, 254–258. doi: 10.1016/j.tics.2008.04.005

Beck, A. T., Steer, R. A., and Hautzinger, M. (2001). *Beck-Depressions-Inventar (BDI): Testhandbuch. 2., überarb. Aufl., 1. Nachdr*. Bern: Huber.

Bekerman, I., Gottlieb, P., and Vaiman, M. (2014). Variations in Eyeball Diameters of the Healthy Adults. *J. Ophthalmol.* 2014, 1–5. doi: 10.1155/2014/503645

Bente, G., Eschenburg, F., and Krämer, N. C. (2007). "Virtual gaze. A pilot study on the effects of computer simulated gaze in avatar-based conversations," in *Virtual reality*. ed R. Shumaker (Berlin: Springer), 185–194. doi: 10.1007/978-3-540-73335-5_21

Bristow, D., Rees, G., and Frith, C. D. (2007). Social interaction modifies neural response to gaze shifts. *Soc. Cogn. Affect. Neurosci.* 2, 52–61. doi: 10.1093/scan/nsl036

Bürkner, P.-C. (2018). Advanced bayesian multilevel modeling with the R Package brms. *R J.* 10, 395–411. doi: 10.32614/RJ-2018-2017

Bürkner, P.-C. (2017). brms?: an R Package for bayesian multilevel models using stan. *J. Stat. Softw.* 80, 1–18. doi: 10.18637/jss.v080.i01

Csibra, G., and Gergely, G. (2009). Natural pedagogy. *Trends Cogn. Sci.* 13, 148–153. doi: 10.1016/j.tics.2009.01.005

Csibra, G., and Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philos. Trans. R. Soc. B Biol. Sci.* 366, 1149–1157. doi: 10.1098/rstb.2010.0319

Emery, N. J. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neurosci. Biobehav. Rev.* 24, 581–604. doi: 10.1016/S0149-7634(00)00025-27

Farroni, T., Mansfield, E. M., Lai, C., and Johnson, M. H. (2003). Infants perceiving and acting on the eyes: tests of an evolutionary hypothesis. *J. Exp. Child Psychol.* 85, 199–212. doi: 10.1016/S0022-0965(03)00022-25

Frischen, A., Bayliss, A. P., and Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychol. Bull.* 133, 694–724. doi: 10.1037/0033-2909.133.4.694

Gelman, A. (2006). Multilevel (Hierarchical) modeling: what it can and cannot do. *Technometrics* 48, 432–435. doi: 10.1198/004017005000000661

Georgescu, A. L., Kuzmanovic, B., Roth, D., Bente, G., and Vogeley, K. (2014). The use of virtual characters to assess and train non-verbal communication in high-functioning autism. *Front. Hum. Neurosci*. 8:807. doi: 10.3389/fnhum.2014.00807

Gibson, J. J., and Pick, A. D. (1963). Perception of another person's looking behavior. *Am. J. Psychol.* 76, 386–394. doi: 10.2307/1419779

Gobel, M. S., Kim, H. S., and Richardson, D. C. (2015). The dual function of social gaze. *Cognition* 136, 359–364. doi: 10.1016/j.cognition.2014.11.040

Gobel, M. S., Tufft, M. R. A., and Richardson, D. C. (2017). Social beliefs and visual attention: how the social relevance of a cue influences spatial orienting. *Cogn. Sci.* 42(Suppl. 1), 161–185. doi: 10.1111/cogs.12529

Gredebäck, G., Astor, K., and Fawcett, C. (2018). Gaze following is not dependent on ostensive cues: a critical test of natural pedagogy. *Child Dev*. 89, 2091–2098. doi: 10.1111/cdev.13026

Hamilton, A. F. de C. (2016). Gazing at me: the importance of social meaning in understanding direct-gaze cues. *Philos. Trans. R. Soc. B Biol. Sci.* 371:20150080. doi: 10.1098/rstb.2015.0080

Hoffman, M. D., and Gelman, A. (2014). The No-U-Turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.* 31, 1593–1623.

Itier, R. J., and Batty, M. (2009). Neural bases of eye and gaze processing: the core of social cognition. *Neurosci. Biobehav. Rev.* 33, 843–863. doi: 10.1016/j.neubiorev.2009.02.004

Jenkins, R., Beaver, J. D., and Calder, A. J. (2006). I thought you were looking at me: direction-specific aftereffects in gaze perception. *Psychol. Sci.* 17, 506–513. doi: 10.1111/j.1467-9280.2006.01736.x

Jording, M., Hartz, A., Bente, G., Schulte-Rüther, M., and Vogeley, K. (2018). The "Social Gaze Space": a taxonomy for gaze-based communication in triadic interactions. *Front. Psychol.* 9:229. doi: 10.3389/fpsyg.2018.00226

Jording, M., Hartz, A., Bente, G., Schulte-Rüther, M., and Vogeley, K. (2019). Inferring interactivity from gaze patterns during triadic person-object-agent interactions. *Front. Psychol*. 10:1913. doi: 10.3389/fpsyg.2019.01913

Kleinke, C. L. (1986). Gaze and eye contact: a research review. *Psychol. Bull.* 100, 78–100. doi: 10.1037//0033-2909.100.1.78

Lehrl, S. (2005). *Manual Zum MWT-B: Mehrfachwahl-Wortschatz-Intelligenztest. 5., unveränd. Aufl*. Balingen: Spitta-Verl.

Lehrl, S., Gallwitz, A., and Blaha, L. (1991). *Theorie und Messung Der Geistigen Leistungsfähigkeit Mit Dem Kurztest KAI: Die Allgemeinen Basisgrößen der Informationsverarbeitung.* Vless: . Fischer Ebersberg.

Lobmaier, J. S., Fischer, M. H., and Schwaninger, A. (2006). Objects capture perceived gaze direction. *Exp. Psychol.* 53, 117–122. doi: 10.1027/1618-3169.53.2.117

Lyyra, P., Wirth, J. H., and Hietanen, J. K. (2017). Are you looking my way? Ostracism widens the cone of gaze. *Q. J. Exp. Psychol*. 70, 1713–1721. doi: 10.1080/17470218.2016.1204327

Mareschal, I., Calder, A. J., and Clifford, C. W. G. (2013). Humans have an expectation that gaze is directed toward them. *Curr. Biol.* 23, 717–721. doi: 10.1016/j.cub.2013.03.030

McElreath, R. (2016). *Statistical Rethinking: a Bayesian course with Examples in R and Stan*. Boca Raton: CRC Press/Taylor & Francis Group.

Morales, M., Mundy, P., and Rojas, J. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behav. Dev.* 21, 373–377. doi: 10.1016/S0163-6383(98)90014-90015

Mundy, P., and Newell, L. (2007). Attention, joint attention, and social cognition. *Curr. Dir. Psychol. Sci.* 16, 269–274. doi: 10.1111/j.1467-8721.2007.00518.x

Paulus, C. (2009). Der Saarbrücker Persönlichkeitsfragebogen SPF(IRI) Zur Messung Von Empathie: Psychometrische Evaluation Der deutschen Version Des Interpersonal Reactivity Index. Available at: http://hdl.handle.net/20.500.11780/3343 (accessed December 05, 2019).

Pfeiffer, U. J., Schilbach, L., Timmermans, B., Kuzmanovic, B., Georgescu, A. L., Bente, G., et al. (2014). Why we interact: on the functional role of the striatum in the subjective experience of social interaction. *NeuroImage* 101, 124–137. doi: 10.1016/j.neuroimage.2014.06.061

Pfeiffer, U. J., Vogeley, K., and Schilbach, L. (2013). From gaze cueing to dual eye-tracking: novel approaches to investigate the neural correlates of gaze in social interaction. *Neurosci. Biobehav. Rev.* 37, 2516–2528. doi: 10.1016/j.neubiorev.2013.07.017

Pönkänen, L. M., Alhoniemi, A., Leppänen, J. M., and Hietanen, J. K. (2011). Does it make a difference if I have an eye contact with you or with your picture? *An ERP study. Soc. Cogn. Affect. Neurosci.* 6, 486–494. doi: 10.1093/scan/nsq068

Qian, Q., Song, M., and Shinomori, K. (2013). Gaze cueing as a function of perceived gaze direction: gaze cueing. *Jpn. Psychol. Res.* 55, 264–272. doi: 10.1111/jpr.12001

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

R Studio Team (2016). *RStudio: Integrated Development for R.* Boston, MA: RStudio, Inc.

Risko, E. F., and Kingstone, A. (2011). Eyes wide shut: implied social presence, eye tracking and attention. *Atten. Percept. Psychophys.* 73, 291–296. doi: 10.3758/s13414-010-0042-41

Senju, A., and Csibra, G. (2008). Gaze Following in human infants depends on communicative signals. *Curr. Biol.* 18, 668–671. doi: 10.1016/j.cub.2008.03.059

Senju, A., Csibra, G., and Johnson, M. H. (2008). Understanding the referential nature of looking: infants' preference for object-directed gaze. *Cognition* 108, 303–319. doi: 10.1016/j.cognition.2008.02.009

Senju, A., and Johnson, M. H. (2009). The eye contact effect: mechanisms and development. *Trends Cogn. Sci.* 13, 127–134. doi: 10.1016/j.tics.2008.11.009

Shepherd, S. V. (2010). Following gaze: gaze-following behavior as a window into social cognition. *Front. Integr. Neurosci.* 4:5. doi: 10.3389/fnint.2010.00005

Stan Development Team (2018). *RStan: the R interface to Stan. R package version 2.14.1*. Available at: http://mc-stan.org/ (accessed December 05, 2019).

Szufnarowska, J., Rohlfing, K. J., Fawcett, C., and Gredebäck, G. (2014). Is ostension any more than attention? *Sci. Rep.* 4:5304. doi: 10.1038/srep05304

Thainimit, S., Alexandre, L. A., and de Almeida, V. M. N. (2013). "Iris surface deformation and normalization," in *2013 13th International Symposium on Communications and Information Technologies (ISCIT)*, (Surat Thani: IEEE), 501–506.

Vernetti, A., Smith, T. J., and Senju, A. (2017). Gaze-contingent reinforcement learning reveals incentive value of social signals in young children and adults. *Proc. R. Soc. B Biol. Sci.* 284, 20162747. doi: 10.1098/rspb.2016.2747

Vogeley, K., and Bente, G. (2010). "Artificial humans": psychology and neuroscience perspectives on embodiment and nonverbal communication. *Neural Netw.* 23, 1077–1090. doi: 10.1016/j.neunet.2010.06.003

von Griinau, M., and Anston, C. (1995). The detection of gaze direction: a stare-in-the-crowd effect. *Perception* 24, 1297–1313. doi: 10.1068/p241297

Walter, H., Adenzato, M., Ciaramidaro, A., Enrici, I., Pia, L., and Bara, B. G. (2004). Understanding intentions in social interaction: the role of the anterior paracingulate cortex. *J. Cogn. Neurosci.* 16, 1854–1863. doi: 10.1162/0898929042947838

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

OPINION
published: 14 January 2020
doi: 10.3389/fpsyg.2019.03005

Body-World Coupling, Sensorimotor Mechanisms, and the Ontogeny of Social Cognition

Daniel Rojas-Líbano and Francisco J. Parada*

Laboratorio de Neurociencia Cognitiva y Social, Facultad de Psicología, Universidad Diego Portales, Santiago, Chile

Keywords: sensorimotor mechanisms, social cognition, mechanistic explanation, ontogeny, 4E-cognition, sensorimotor coupling

Edited by:
Peter König,
University of Osnabrück, Germany

Reviewed by:
Miguel Aguilera,
University of the Basque Country, Spain
Valentina Fantasia,
Max Planck Institute for Human Development, Germany

****Correspondence:***
Daniel Rojas-Líbano
daniel.rojas@udp.cl

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *08 February 2019*
Accepted: *18 December 2019*
Published: *14 January 2020*

Citation:
Rojas-Líbano D and Parada FJ (2020) Body-World Coupling, Sensorimotor Mechanisms, and the Ontogeny of Social Cognition. Front. Psychol. 10:3005. doi: 10.3389/fpsyg.2019.03005

When closely examined, several biological mechanisms reveal themselves as implementing a physical and dynamical two-way link or coupling between the organism and the world. In these cases, some mechanisms' components can either physically cross the body-world boundary or are brought by the organism's motor actions onto specific sensory surfaces. As with any biological phenomenon, the historical contingencies of these sensorimotor activities generate plastic changes within the organism, that in turn determine its capacities at any given time. Body-world coupling instances are evident in examples that we will describe later, such as breathing, sensori-motor activities, and others. In the present piece, we attempt to position social cognitive phenomena as the result of the mechanisms involved in the organism's coupling history with its world. This coupling constitutes one of the cornerstones of the so called 4E approach to cognition (Newen et al., 2018), from which we will also draw concepts and distinctions in our effort to relate coupling mechanisms with social phenomena. Even though reviewing the 4E approach to cognition escapes the scope of the present piece, we can briefly state that the 4E cognition framework wants to bring multiple approaches together under a sole emblem. It understands cognition as a natural phenomenon, *embodied* in the biophysics of the body which is *embedded* both phylo- and ontogenically into the animal's ecological niche. To the 4E approach, cognition is also *opportunistic* and *promiscuous* as can be *extended* toward the world with objects both material (e.g., technology) and conceptual (e.g., institutions). Finally, the 4E approach thinks cognition as intended for action in an ongoing interactional sense-making process; an *enactive* phenomenon. The 4E cognition framework owes its current form to several landmark work such as the "enactive approach" (Varela et al., 2017), the "distributed cognition branch of cognitive science" (Flor and Hutchins, 1991; Hutchins, 1995), and the "extended mind" proposal (Clark and Chalmers, 1998), among others.

Despite decades of conceptual development of the 4E approach and its diverse subfields, there are many questions regarding its particular implications for neuroscience (e.g., how can neuroscientists can actually implement the 4E approach directly into their research agendas? Is one-person neuroscience necessary?, etc.) (Di Paolo and De Jaegher, 2012; Willems and Francken, 2012). As experimental neuroscientists interested in the interactional nature of cognition, we would like to extract the mechanistic implications of the 4E approach: components, activities, and processes (What?, How?, When?), their context (When?, How?) and their weights (How important?). Epistemologically, we concur with the view that conceives mechanisms as models of the phenomena to explain and consider the building of mechanistic models a fundamental explanatory aim of neuroscience (Craver, 2007). Without a mechanistic picture of the ways in which the 4Es constitute and/or affect cognitive processes, we are left with few tools to further empirical research.

We start by considering relevant distinctions provided by De Jaegher et al. (2010), where *constitutive*, *enabling*, and *contextual* factors can be identified as the "set of circumstances" which are phenomena themselves. A *contextual* factor modulates the phenomenon, whilst an *enabling* one is necessary for the phenomenon to occur. Finally, *constitutive* factors are processes, parts, and/or pieces that produce the phenomenon itself. What happens if we add a dynamic and mechanistic framework to the De Jaegher, Di Paolo, and Gallagher's proposal? The phenomenon to explain -at any scale (from action potentials to social interaction)- can be understood as the result of the dynamic operation of one or more mechanisms. Such mechanisms comprise components, their activities and the processes in which they participate, whose structural and functional organization in certain conditions produce the phenomenon (Bechtel and Abrahamsen, 2005; Craver, 2007). Thus, we suggest that *constitutive* factors are processes that can be composed of different components of a mechanism under consideration at different moments of time. Examples of components participating in a constitutive fashion are ion channels, for the phenomenon of the action potential, and participating agents for social interaction. In contrast, *contextual* and *enabling* factors are better understood here as elements which interact with mechanisms' components and can change its operation regime. Examples of enabling factors are the existence of ionic gradients across the membrane, for the action potential, and the alertness level of a participant, for social interaction. Examples of contextual factors are, a specific ion channel type for the action potential, and a given environmental setting, for social interaction. It is important to note here that the constitutive, enabling, or contextual quality of a given factor it is not fixed, but can change throughout the organism's ontogeny or history of structural change.

We think our mechanistic view is compatible with the original proposal of De Jaegher et al. (2010). In what follows, we consider the above mentioned points in some detail. We start by examining different mechanisms of body-world coupling, to then propose ways to extend this viewpoint into social-cognitive phenomena, considering the organism's ontogeny.

BODY-WORLD COUPLING

Active Coupling Through Sensorimotor Activities

An example of body-world coupling is represented by an animal's sensory-motor activities. In situations where the sensory processes are important for the organism, there is usually a profound interplay between the animal's actions and the operation of its sensors (Rojas-Líbano et al., 2014). This is evident in motor actions associated with sensory sampling of the environment: touching, sniffing, echolocating, whisking, visual scanning. These actions allow the animal to bring stimuli to sensory surfaces. In most of these cases, stimuli sampling takes place in the wider context of adaptive and context-sensitive behavior. The animal actively moves its sensory systems to make decisions about navigation, small displacements, further explorations, language actions, and the like (Ganguly and Kleinfeld, 2004; Hayhoe and Ballard, 2005; Rojas-Líbano and Kay, 2012; Clark, 2013; Arce-McShane et al., 2016).

The appropriate interplay or coordination between motor actions and sensory activations requires the participation of certain components of the world in the sensory-motor mechanism. Therefore, cognitive activities involving any type of movement will demand some environmental components to become participants of a mechanism (i.e., a transiently constitutive factor). If we manipulate world conditions that interfere with this loop, we can potentially destroy the organism's coupling in the sense that we decrease its ability to interact coherently with its world. Examples are everywhere. Sniffs manipulate the number of odor molecules drawn onto the olfactory epithelium, as well as the rate (i.e., flow) at which those molecules travel through the nose (Rojas-Líbano and Kay, 2012). Tactile (e.g., whisker, finger) movements are coordinated with body movements and control the spatiotemporal frequency at which mechanical stimuli contact the skin cutaneous receptors (Kurnikova et al., 2017). Eye/head/body movements effectively displace the photoreceptor surface so as to receive photons coming from specific objects from the visual scene (Schroeder et al., 2010), and mechanisms such as the accommodation reflex modify the amount and direction of light that reaches the retina, via the modification of pupil size and lens width (Michael-Titus et al., 2010). All these motor activities *manipulate* world components and -through this manipulation- cause changes onto sensory surfaces (**Figures 1A,B**). Thus, world components continuously move back and forth from participating in processes contextual or enabling to constitutive factors for a given point in time and a given sensorimotor act.

Other Examples of Coupling

Some mechanisms are part of the basic autonomy of a living being and can be independent of active volitional control. There are many examples, such as coupling through circadian rhythms or, at the cellular scale, through membrane potential maintenance, nutrient exchange, and structural interactions with the extracellular matrix. However, for the sake of simplicity, let us specifically focus on mammalian breathing as a non-sensorimotor example of a mechanism that allows an organism to functionally couple with its world. We know a fairly good deal of the neural mechanisms that implement breathing in mammals (Feldman and Del Negro, 2006). In this process, the animal actively exchanges components with its world, specifically air volumes with different amounts of oxygen and carbon dioxide. Neurons in the brainstem periodically fire impulses that eventually send activity down the phrenic nerve, delivering acetylcholine onto the muscle cells of the diaphragm. The diaphragm then contracts, expanding the thoracic cavity and increasing lung volume. This expansion draws air from the organism's surroundings into the lungs. Finally, the diaphragm relaxes, pushing air from inside the lungs back to the exterior of the animal's body. Accompanying the volume exchange there is a substance exchange: inspired air is more enriched in oxygen than expired air, which in turn is more enriched in carbon dioxide. At a molecular scale, we can conceive the mechanism as a continuous exchange of molecules. From an outside reservoir enriched in oxygen molecules, the organism draws oxygen inside and pushes

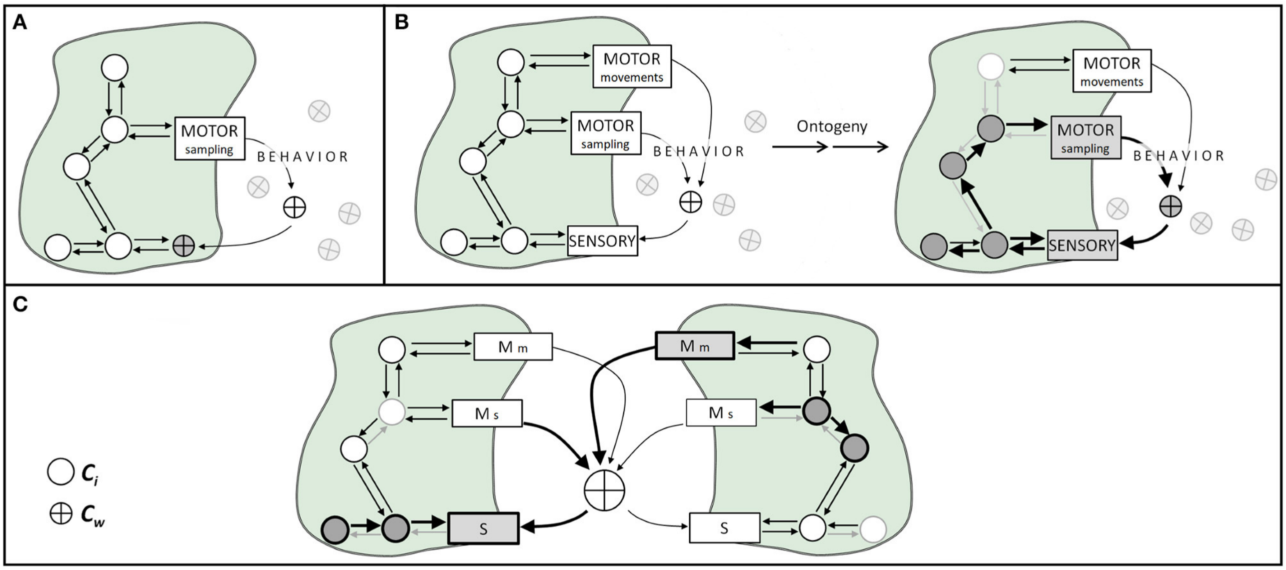


FIGURE 1 | Illustration of body-world coupling, sensorimotor mechanisms, and the ontogeny of social cognition. Circles represent mechanisms' components internal to the organism (C_i), and crossed circles depict world components (C_w). Arrows represent causal effects between components. **(A)** In non-sensorimotor body-world couplings, an organism's motor activities capture world's components and makes them interact with body components. **(B)** In sensorimotor body-world couplings, through active motor sampling activities (e.g., sniffing, touching, fixating) the organism dynamically brings world components onto sensory surfaces. During ontogeny, the occurrence (or not) of specific sensorimotor activities produces plastic changes in body components, represented here as weighted arrows and circles. **(C)** Social cognition as a process grounded in sensorimotor coupling. World components relevant to sensorimotor coupling could well be stimuli produced by another organism, such as physical stimuli resulting from communication processes, and different agents coupled with shared world components can lead to social cognitive phenomena. This kind of sensorimotor coupling might entail different plastic processes within each organism (represented as different weighted arrows inside each agent). S, sensory; Mm, motor actions for movement; Ms, motor actions for sampling.

out carbon dioxide. This mechanism operates as long as the animal preserves its biological autonomy.

Now, consider what happens when we intervene on the external side. Lowering the air oxygen concentration causes a decrease in blood oxygen, which in turn activates peripheral and central chemoreceptor neurons (Teppema and Dahan, 2010). The activation of the latter triggers an increase in drive to the diaphragm, resulting in stronger, and more frequent breathing cycles. Something similar happens if we prevent molecules from crossing the boundary, say by occluding the airway. This indicates that by manipulating the external state of affairs, and/or by preventing physical exchanges across the body-world boundary, we causally intervene in the mechanism. We propose that this is a feature of mechanisms that couple body and world. It is also trivially true that several manipulations of the external conditions can causally affect the body, such as when the body is hit, for example, by a heavy object. But in those cases the world component involved was not implicated in a regular mechanism with the organism.

ONTOGENY, SOCIAL COGNITION, AND BODY-WORLD COUPLING MECHANISMS

In the cases described above, and in many others, what we see is a physiological mechanism that contains -as part of its regular components- some element(s) of the world. By altering either internal or external components, we alter the mechanism operation (**Figures 1A,B**).

Let ***M*** be a (neuro)physiological mechanism (e.g., respiration, sensorimotor operations, circadian rhythms) containing internal components ***Ci*** which normally interact with some world's components ***Cw*** (any processes and/or entities, whether living or not, present outside the organism's physical body). Traditionally, it is conceived that the operation of ***M*** depends on ***Ci*** alone. However, for relevant biological phenomena, such as respiration or sensorimotor activities, ***Cw*** are mechanism components, participating in the resulting processes, and therefore we think is useful to regard them as constitutive[1]. Likewise, other ***Cw*** would be enabling and/or contextual, depending on the phenomena under consideration. Considering ***Cw*** as constitutive and/or enabling elements of a given ***M***, we can further state that many organizational principles of the brain -generated from multiple operating mechanisms- will be much better explained by incorporating their relationship to the world (Clark and Chalmers, 1998; Cosmelli and Thompson, 2010; Parada and Rossi, 2018).

We could also say that the operation of a given ***M*** will depend on the organism's past and current temporospatial

---

[1] We follow Craver (2007) in using manipulability as a criterion for recognizing mechanisms' components. Briefly stated, if interventions on the mechanism as a whole are accompanied by changes of a potential component, and if interventions on the component produce, in turn, changes in the mechanism, then the component under consideration is a mechanism's component.

contingencies (i.e., both ***Ci*** and ***Cw***). A key notion here is that biological mechanisms are not timeless laws, but *historically contingent processes* (Craver, 2007). Consider, as an example, the mechanisms of neural plasticity. It has been shown that present neuronal properties -both structural and functional[2]- are dependent on the neuron's previous interactions with its immediate environment (Rose and Rankin, 2001; Bailey et al., 2015; Andersen et al., 2017; Schulz and Lane, 2017). Importantly, this is not a special feature of neurons, but a general biological phenomenon. The actual state and capacities of any organism are activity- and ontogeny-dependent, and are always intertwined with the environment in which ontogeny takes place (Stagg et al., 2011; Kelly et al., 2012; Ganguly and Poo, 2013; Sale et al., 2014; Fields, 2015). Social-cognitive phenomena can be conceived, within this framework, as interactions occurring through the sharing of some ***Cw*** between the agents engaged in it (**Figure 1C**).

Taking into account the dependence on history of biological mechanisms, it is particularly relevant to distinguish the role of ***Cw*** at different moments along ontogeny. At different moments, the weight of a ***Cw*** could play a role as a *constitutive*, *enabling*, or *contextual* factors in a given phenomenon. For example, the case of behavioral habituation shows that, under sustained interactions, responses to the same ***Cw*** can decrease drastically, turning a ***Cw*** stimulus from a once-constitutive element to a mere contextual perturbation (Brunelli et al., 1976). In what follows, we use these ideas to propose a link between ontogenic mechanisms of body-world coupling and social interactions.

Social interaction starts very early during development, from prenatal experiences to turn-taking in babies to early verbalizations in infants (Siddiqui and Hägglöf, 2000; Kugiumutzakis, 2017; Quigley et al., 2017). From the point of view of mechanisms of body-world coupling, these developmental changes correspond to an increment in the allowed complexity of sensorimotor interactions. Mechanistically, increased sensorimotor complexity can be reached by reducing the sensorimotor contingencies' dimensionality, using both history of interactions and sensorimotor function. This is the organism's current morphological shape, as a product of previous body-world couplings in time, affords more complex actions contained in appropriate ecological niches. A now-classic example is the theoretical (Smith et al., 1999) and empirical (Smith and Thelen, 2003) dynamical systems account of the A-not-B error in infants (Piaget, 1962). Briefly, the processes underlying the perseverative reaching seen in the A-not-B error are not only continuously tied to the infant's sensorimotor system but also to her history of interactions (Spencer et al., 2011). From our perspective, evidence from animal models suggests a constitutive role of external factors such as maternal state during gestation (Kofman, 2002), maternal care/physical contact (Cancedda et al., 2004; Sale et al., 2004), as well as overall environmental conditions (Cai et al., 2009). Similar effects have been reported in humans; social, cultural, and/or physical environmental conditions in earlier developmental stages might bias -or even shape- bio-psycho-social trajectories (Guzzetta et al., 2009; Bowers and Yehuda, 2016; McEwen, 2017). Later in life, most of these factors can become enabling and/or contextual.

A more speculative example -directly related to social cognition- could be found in language; a higher-level cognitive phenomenon profoundly sensitive to ontogenic changes (Peña et al., 2003; Dehaene-Lambertz et al., 2008; Mampe et al., 2009; Mahmoudzadeh et al., 2013; Werker and Gervain, 2013; Werker and Hensch, 2015). The available evidence indicates that human auditory learning starts in the third trimester of gestation (Shahidullah and Hepper, 1994; Hepper, 1996). We further interpret this evidence as suggesting a constitutive role for prenatal listening experiences (***Cw***) in the specific physiological and developmental trajectory that gives rise to speech processing brain structures (***Ci***) (Wermke and Friederici, 2004). Between the 8th and 10th month of age, this body-world coupling begins its consolidation, allowing infants to extract statistical regularities (Saffran et al., 1996), which we conceive as a dimensionality reduction of the complex linguistic world (Werker and Tees, 1984; Maurer and Werker, 2014)[3]. Following our interpretation of these data, listening experiences and verbal interactions (***Cw***) become contextual factors after the 10th month of age (Werker and Curtin, 2005; Werker and Hensch, 2015). We further speculate that such change, from constitutive to contextual, illustrates the dimensionality reduction required for the appearance of more complex sensorimotor operations, such as actively seeking learning opportunities, maximizing informative interactions, and the beginning of adult-like social interactions (Begus et al., 2016). We still lack both data and tools to appropriately model the role, weight, and influence of external factors (from physical interplay to social interactions to processes unfolding from them) in the emergence of social-cognitive functioning and the overall biophysics of human experience.

CLOSING REMARKS

The present opinion piece seeks to facilitate a mechanistic approximation to multi-level phenomena, grounding social cognition, and social interaction into time-dependent functional and structural components and their interplay; a goal for the 4E approach to cognition. Furthermore, it points to the need of modeling, through experimental manipulations, the weight and influence of both internal [i.e., (neuro)physiological] and external (i.e., objects, processes, other people) components at a given developmental period. This modeling can be achieved through tools derived from network science and/or machine learning techniques (Vespignani, 2011; Boonstra et al., 2015; Sekara et al., 2016; Shine et al., 2016; Avena-Koenigsberger et al., 2017; Aguilera, 2018; Parada and Rossi, 2018). Furthermore, implementing scalable experimental paradigms (Parada, 2018; Matusz et al., 2019; Shamay-Tsoory and Mendelsohn, 2019) and generating novel hypotheses of interacting brain/body systems functioning during natural cognition (De Jaegher et al., 2010, 2016; Di Paolo and De Jaegher, 2012; Gramann et al., 2014; Ladouce et al., 2017; Parada, 2018; Parada and Rossi, 2018) are among the most outstanding challenges for the 4E-cognition

---

[2] And hence of the networks in which the neuron participates.

[3] This is also seen in other aspects of perceptual development (Scott et al., 2007).

research program. We believe that the incorporation of a mechanistic framework facilitates meeting those challenges and advancing a deeper understanding of cognitive phenomena, social, and otherwise.

AUTHOR CONTRIBUTIONS

DR-L and FP conceptualized the present work and wrote the current version for publication.

FUNDING

The present work was supported by Comisión Nacional de Investigación Científica y Tecnológica (CONICYT, Chile), through FONDECYT Iniciación en Investigación Grant No. 11180620 awarded to FP. FP was also funded from FONDECYT Regular Grant No. 1170292. Both DR-L and FP receive funding from FONDECYT Regular Grant No. 1190610. Funding sources (CONICYT, Chile) had no role in the production and development of this manuscript. DR-L receives funding from FONDECYT 1180932.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Ismael Palacios and Dr. Alejandra Rossi for insightful comments and constant meaningful conversations during the writing of this work. We would also like to thank the reviewers for their deeply constructive comments and suggestions during the review process.

REFERENCES

Aguilera, M. (2018). Rhythms of the collective brain: metastable synchronization and cross-scale interactions in connected multitudes. *Complexity* 2018:4212509. doi: 10.1155/2018/4212509

Andersen, N., Krauth, N., and Nabavi, S. (2017). Hebbian plasticity *in vivo*: relevance and induction. *Curr. Opin. Neurobiol.* 45, 188–192. doi: 10.1016/j.conb.2017.06.001

Arce-McShane, F. I., Ross, C. F., Takahashi, K., Sessle, B. J., and Hatsopoulos, N. G. (2016). Primary motor and sensory cortical areas communicate via spatiotemporally coordinated networks at multiple frequencies. *Proc. Natl. Acad. Sci. U.S.A.* 113, 5083–5088. doi: 10.1073/pnas.1600788113

Avena-Koenigsberger, A., Misic, B., and Sporns, O. (2017). Communication dynamics in complex brain networks. *Nat. Rev. Neurosci.* 19, 17–33. doi: 10.1038/nrn.2017.149

Bailey, C. H., Kandel, E. R., and Harris, K. M. (2015). Structural components of synaptic plasticity and memory consolidation. *Cold Spring Harb. Perspect. Biol.* 7:a021758. doi: 10.1101/cshperspect.a021758

Bechtel, W., and Abrahamsen, A. (2005). Explanation: a mechanist alternative. *Stud. Hist. Philos. Biol. Biomed. Sci.* 36, 421–441. doi: 10.1016/j.shpsc.2005.03.010

Begus, K., Gliga, T., and Southgate, V. (2016). Infants' preferences for native speakers are associated with an expectation of information. *Proc. Natl. Acad. Sci. U.S.A.* 113, 12397–12402. doi: 10.1073/pnas.1603261113

Boonstra, T. W., Danna-Dos-Santos, A., Xie, H.-B., Roerdink, M., Stins, J. F., and Breakspear, M. (2015). Muscle networks: Connectivity analysis of EMG activity during postural control. *Sci. Rep.* 5:17830. doi: 10.1038/srep17830

Bowers, M. E., and Yehuda, R. (2016). Intergenerational transmission of stress in humans. *Neuropsychopharmacology* 41, 232–244. doi: 10.1038/npp.2015.247

Brunelli, M., Castellucci, V., and Kandel, E. R. (1976). Synaptic facilitation and behavioral sensitization in Aplysia: possible role of serotonin and cyclic AMP. *Science* 194, 1178–1181. doi: 10.1126/science.186870

Cai, R., Guo, F., Zhang, J., Xu, J., Cui, Y., and Sun, X. (2009). Environmental enrichment improves behavioral performance and auditory spatial representation of primary auditory cortical neurons in rat. *Neurobiol. Learn. Memory* 91, 366–376. doi: 10.1016/j.nlm.2009.01.005

Cancedda, L., Putignano, E., Sale, A., Viegi, A., Berardi, N., and Maffei, L. (2004). Acceleration of visual system development by environmental enrichment. *J. Neurosci.* 24, 4840–4848. doi: 10.1523/JNEUROSCI.0845-04.2004

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477

Clark, A., and Chalmers, D. (1998). The extended mind. *Analysis* 58, 7–19. doi: 10.1093/analys/58.1.7

Cosmelli, D., and Thompson, E. (2010). "Embodiment or envatment? Reflections on the bodily basis of consciousness," in *Enaction: Toward a New Paradigm for Cognitive Science,* eds J. Stewart, O. Gapenne, and E. Di Paolo (Cambridge, MA: The MIT Press), 361–385.

Craver, C. F. (2007). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. New York, NY: Clarendon Press.

De Jaegher, H., Di Paolo, E., and Adolphs, R. (2016). What does the interactive brain hypothesis mean for social neuroscience? A dialogue. *Philos. Trans. R. Soc. Lond., B, Biol. Sci. B Biol. Sci.* 371:20150379. doi: 10.1098/rstb.2015.0379

De Jaegher, H., Di Paolo, E., and Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends Cogn. Sci.* 14, 441–447. doi: 10.1016/j.tics.2010.06.009

Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., and Dehaene, S. (2008). How does early brain organization promote language acquisition in humans? *Euro. Rev.* 16, 399–411. doi: 10.1017/S1062798708000513

Di Paolo, E., and De Jaegher, H. (2012). The interactive brain hypothesis. *Front. Hum. Neurosci.* 6:163. doi: 10.3389/fnhum.2012.00163

Feldman, J. L., and Del Negro, C. A. (2006). Looking for inspiration: new perspectives on respiratory rhythm. *Nat. Rev. Neurosci.* 7, 232–242. doi: 10.1038/nrn1871

Fields, R. D. (2015). A new mechanism of nervous system plasticity: activity-dependent myelination. *Nat. Rev. Neurosci.* 16, 756–767. doi: 10.1038/nrn4023

Flor, N. V., and Hutchins, E. L. (1991). "A case study of team programming during perfective software maintenance," in *Empirical Studies of Programmers: Fourth Workshop* eds J. Koenemann-Belliveau, T. Glenn Moher, and S. P. Robertson (Norwood, NJ: Intellect Books), 36.

Ganguly, K., and Kleinfeld, D. (2004). Goal-directed whisking increases phase-locking between vibrissa movement and electrical activity in primary sensory cortex in rat. *Proc. Natl. Acad. Sci. U.S.A.* 101, 12348–12353. doi: 10.1073/pnas.0308470101

Ganguly, K., and Poo, M.-M. (2013). Activity-dependent neural plasticity from bench to bedside. *Neuron* 80, 729–741. doi: 10.1016/j.neuron.2013.10.028

Gramann, K., Ferris, D. P., Gwin, J., and Makeig, S. (2014). Imaging natural cognition in action. *Int. J. Psychophysiol.* 91, 22–29. doi: 10.1016/j.ijpsycho.2013.09.003

Guzzetta, A., Baldini, S., Bancale, A., Baroncelli, L., Ciucci, F., Ghirri, P., et al. (2009). Massage accelerates brain development and the maturation of visual function. *J. Neurosci.* 29, 6042–6051. doi: 10.1523/JNEUROSCI.5548-08.2009

Hayhoe, M., and Ballard, D. (2005). Eye movements in natural behavior. *Trends Cogn. Sci.* 9, 188–194. doi: 10.1016/j.tics.2005.02.009

Hepper, P. G. (1996). Fetal memory: does it exist? What does it do? *Acta Paediatr.* 85, 16–20. doi: 10.1111/j.1651-2227.1996.tb14272.x

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.

Kelly, S. A., Panhuis, T. M., and Stoehr, A. M. (2012). Phenotypic plasticity: molecular mechanisms and adaptive significance. *Compr. Physiol.* 2, 1417–1439. doi: 10.1002/cphy.c110008

Kofman, O. (2002). The role of prenatal stress in the etiology of developmental behavioural disorders. *Neurosci. Biobehav. Rev.* 26, 457–470. doi: 10.1016/S0149-7634(02)00015-5

Kugiumutzakis, G. (2017). "Intersubjective vocal imitation in early mother-infant interaction," in *New Perspectives in Early Communicative Development* eds J. Nadel and L. Camaioni (New York, NY: Routledge), 23–47.

Kurnikova, A., Moore, J. D., Liao, S.-M., Deschênes, M., and Kleinfeld, D. (2017). Coordination of orofacial motor actions into exploratory behavior by rat. *Curr. Biol. CB*, 27, 688–696. doi: 10.1016/j.cub.2017.01.013

Ladouce, S., Donaldson, D. I., Dudchenko, P. A., and Letswaart, M. (2017). Understanding minds in real-world environments: toward a mobile cognition approach. *Front. Human Neurosci.* 10:694. doi: 10.3389/fnhum.2016.00694

Mahmoudzadeh, M., Dehaene-Lambertz, G., Fournier, M., Kongolo, G., Goudjil, S., Dubois, J., et al. (2013). Syllabic discrimination in premature human infants prior to complete formation of cortical layers. *Proc. Natl. Acad. Sci. U.S.A.* 110, 4846–4851. doi: 10.1073/pnas.1212220110

Mampe, B., Friederici, A. D., Christophe, A., and Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Curr. Biol. CB* 19, 1994–1997. doi: 10.1016/j.cub.2009.09.064

Matusz, P. J., Dikker, S., Huth, A. G., and Perrodin, C. (2019). Are we ready for real-world neuroscience? *J. Cogn. Neurosci.* 31, 327–338. doi: 10.1162/jocn_e_01276

Maurer, D., and Werker, J. F. (2014). Perceptual narrowing during infancy: a comparison of language and faces. *Dev. Psychobiol.* 56, 154–178. doi: 10.1002/dev.21177

McEwen, B. S. (2017). Allostasis and the epigenetics of brain and body health over the life course: the brain on stress. *JAMA Psychiatry.* 74, 551–552. doi: 10.1001/jamapsychiatry.2017.0270

Michael-Titus, A., Revest, P., and Shortland, P. (2010). "7 - the visual system," in *The Nervous System, 2nd Edn,* eds A. Michael-Titus, P. Revest, and P. Shortland (London, UK: Churchill Livingstone), 121–140.

Newen, A., De Bruin, L., and Gallagher, S (eds.). (2018). *The Oxford handbook of 4E Cognition*. Croydon: Oxford University Press.

Parada, F. J. (2018). Understanding natural cognition in everyday settings: 3 pressing challenges. *Front. Human Neurosci.* 12:386. doi: 10.3389/fnhum.2018.00386

Parada, F. J., and Rossi, A. (2018). If neuroscience needs behavior what does psychology need? *Front. Psychol.* 9:433. doi: 10.3389/fpsyg.2018.00433

Peña, M., Maki, A., Kovačić, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., et al. (2003). Sounds and silence: an optical topography study of language recognition at birth. *Proc. Natl. Acad. Sci. U.S.A.* 100, 11702–11705. doi: 10.1073/pnas.1934290100

Piaget, J. (1962). *Play, Dreams and Imitation in Childhood.* New York: Norton.

Quigley, K. M., Moore, G. A., Propper, C. B., Goldman, B. D., and Cox, M. J. (2017). Vagal regulation in breastfeeding infants and their mothers. *Child Dev.* 88, 919–933. doi: 10.1111/cdev.12641

Rojas-Líbano, D., Frederick, D. E., Egaña, J. I., and Kay, L. M. (2014). The olfactory bulb theta rhythm follows all frequencies of diaphragmatic respiration in the freely behaving rat. *Front. Behav. Neurosci.* 8:214. doi: 10.3389/fnbeh.2014.00214

Rojas-Líbano, D., and Kay, L. M. (2012). Interplay between sniffing and odorant sorptive properties in the rat. *J. Neurosci.* 32, 15577–15589. doi: 10.1523/JNEUROSCI.1464-12.2012

Rose, J. K., and Rankin, C. H. (2001). Analyses of habituation in Caenorhabditis elegans. *Learn. Memory* 8, 63–69. doi: 10.1101/lm.37801

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926–1928. doi: 10.1126/science.274.5294.1926

Sale, A., Berardi, N., and Maffei, L. (2014). Environment and brain plasticity: towards an endogenous pharmacotherapy. *Physiol. Rev.* 94, 189–234. doi: 10.1152/physrev.00036.2012

Sale, A., Putignano, E., Cancedda, L., Landi, S., Cirulli, F., Berardi, N., et al. (2004). Enriched environment and acceleration of visual system development. *Neuropharmacology* 47, 649–660. doi: 10.1016/j.neuropharm.2004.07.008

Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., and Lakatos, P. (2010). Dynamics of active sensing and perceptual selection. *Curr. Opin. Neurobiol.* 20, 172–176. doi: 10.1016/j.conb.2010.02.010

Schulz, D. J., and Lane, B. J. (2017). Homeostatic plasticity of excitability in crustacean central pattern generator networks. *Curr. Opin. Neurobiol.* 43, 7–14. doi: 10.1016/j.conb.2016.09.015

Scott, L. S., Pascalis, O., and Nelson, C. A. (2007). A domain-general theory of the development of perceptual discrimination. *Curr. Dir. Psychol. Sci.* 16, 197–201. doi: 10.1111/j.1467-8721.2007.00503.x

Sekara, V., Stopczynski, A., and Lehmann, S. (2016). Fundamental structures of dynamic social networks. *Proc. Natl. Acad. Sci. U.S.A.* 113, 9977–9982. doi: 10.1073/pnas.1602803113

Shahidullah, S., and Hepper, P. G. (1994). Frequency discrimination by the fetus. *Early Hum. Dev.* 36, 13–26. doi: 10.1016/0378-3782(94)90029-9

Shamay-Tsoory, S. G., and Mendelsohn, A. (2019). Real-life neuroscience: an ecological approach to brain and behavior research. *Perspect. Psychol. Sci.* 14, 841–859. doi: 10.1177/1745691619856350

Shine, J. M., Bissett, P. G., Bell, P. T., Koyejo, O., Balsters, J. H., Gorgolewski, K. J., et al. (2016). The dynamics of functional brain networks: integrated network states during cognitive task performance. *Neuron* 92, 544–554. doi: 10.1016/j.neuron.2016.09.018

Siddiqui, A., and Hägglöf, B. (2000). Does maternal prenatal attachment predict postnatal mother–infant interaction? *Early Hum. Dev.* 59, 13–25. doi: 10.1016/s0378-3782(00)00076-1

Smith, L. B., and Thelen, E. (2003). Development as a dynamic system. *Trends Cogn. Sci.* 7, 343–348. doi: 10.1016/s1364-6613(03)00156-6

Smith, L. B., Thelen, E., Titzer, R., and McLin, D. (1999). Knowing in the context of acting: the task dynamics of the A-not-B error. *Psychol. Rev.* 106:235. doi: 10.1037/0033-295x.106.2.235

Spencer, J. P., Perone, S., and Buss, A. T. (2011). Twenty years and going strong: a dynamic systems revolution in motor and cognitive development. *Child Dev. Perspect.* 5, 260–266. doi: 10.1111/j.1750-8606.2011.00194.x

Stagg, N. J., Mata, H. P., Ibrahim, M. M., Henriksen, E. J., Porreca, F., Vanderah, T. W., et al. (2011). Regular exercise reverses sensory hypersensitivity in a rat neuropathic pain model: role of endogenous opioids. *Anesthesiology* 114, 940–948. doi: 10.1097/ALN.0b013e318210f880

Teppema, L. J., and Dahan, A. (2010). The ventilatory response to hypoxia in mammals: mechanisms, measurement, and analysis. *Physiol. Rev.* 90, 675–754. doi: 10.1152/physrev.00012.2009

Varela, F. J., Thompson, E., and Rosch, E. (2017). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.

Vespignani, A. (2011). Modelling dynamical processes in complex socio-technical systems. *Nat. Phys.* 8, 32–39. doi: 10.1038/nphys2160

Werker, J. F., and Curtin, S. (2005). PRIMIR: a developmental framework of infant speech processing. *Lang. Learn. Dev.* 1, 197–234. doi: 10.1080/15475441.2005.9684216

Werker, J. F., and Gervain, J. (2013). "Speech perception in infancy: a foundation for language acquisition," in *The Oxford Handbook of Developmental Psychology*, Vol. 1, ed P. David Zelazo (New York, NY: Oxford University Press), 909–925.

Werker, J. F., and Hensch, T. K. (2015). Critical periods in speech perception: new directions. *Annu. Rev. Psychol.* 66, 173–196. doi: 10.1146/annurev-psych-010814-015104

Werker, J. F., and Tees, R. C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behav. Dev.* 7, 49–63. doi: 10.1016/S0163-6383(84)80022-3

Wermke, K., and Friederici, A. D. (2004). Developmental changes of infant cries–the evolution of complex vocalizations. *Behav. Brain Sci.* 27, 474–475. doi: 10.1017/S0140525X04390102

Willems, R. M., and Francken, J. C. (2012). Embodied cognition: taking the next step. *Front. Psychol.* 3:582. doi: 10.3389/fpsyg.2012.00582

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 04 February 2020
doi: 10.3389/fpsyg.2020.00079

Audiovisual Integration During Joint Action: No Effects for Motion Discrimination and Temporal Order Judgment Tasks

Basil Wahn*, **Jill A. Dosso and Alan Kingstone**

Department of Psychology, University of British Columbia, Vancouver, BC, Canada

In daily life, humans constantly process information from multiple sensory modalities (e.g., visual and auditory). Information across sensory modalities may (or may not) be combined to form the perception of a single event via the process of multisensory integration. Recent research has suggested that performing a spatial crossmodal congruency task jointly with a partner affects multisensory integration. To date, it has not been investigated whether multisensory integration in other crossmodal tasks is also affected by performing a task jointly. To address this point, we investigated whether joint task performance also affects perceptual judgments in a crossmodal motion discrimination task and a temporal order judgment task. In both tasks, pairs of participants were presented with auditory and visual stimuli that might or might not be perceived as belonging to a single event. Each participant in a pair was required to respond to stimuli from one sensory modality only (e.g., visual stimuli only). Participants performed both individual and joint conditions. Replicating earlier multisensory integration effects, we found that participants' perceptual judgments were significantly affected by stimuli in the other modality for both tasks. However, we did not find that performing a task jointly modulated these crossmodal effects. Taking this together with earlier findings, we suggest that joint task performance affects crossmodal results in a manner dependent on how these effects are quantified (i.e., via responses time or accuracy) and the specific task demands (i.e., whether tasks require processing stimuli in terms of location, motion, or timing).

Keywords: multisensory integration, social cognition, joint action, motion, temporal, task co-performance

Edited by:
Andreas K. Engel,
University Medical Center
Hamburg-Eppendorf, Germany

Reviewed by:
Valerio Santangelo,
University of Perugia, Italy
Valter Prpic,
De Montfort University,
United Kingdom

Correspondence:
Basil Wahn
basilwahn@gmail.com

Specialty section:
This article was submitted to
Cognitive Science,
a section of the journal
Frontiers in Psychology

Received: *19 July 2019*
Accepted: *13 January 2020*
Published: *04 February 2020*

Citation:
Wahn B, Dosso JA and Kingstone A
(2020) Audiovisual Integration During
Joint Action: No Effects for Motion
Discrimination and Temporal Order
Judgment Tasks.
Front. Psychol. 11:79.
doi: 10.3389/fpsyg.2020.00079

1. INTRODUCTION

Humans constantly process information from several sensory modalities (e.g., touch, vision, audition). This information may (or may not) be combined to form a unitary percept via the process of multisensory integration. Previous research has investigated several factors that could affect this integration process ranging from where and when the stimuli occur (Meredith et al., 1987; Meredith and Stein, 1996; Guski and Troje, 2003; Holmes and Spence, 2005) to the attentional demands placed on the observer (Bertelson et al., 2000; Vroomen et al., 2001; Alsius et al., 2005; Santangelo and Spence, 2008, 2009; Santangelo and Macaluso, 2012; Vercillo and Gori, 2015; Wahn and König, 2015a,b, 2016; Wahn et al., 2017b).

To date, however, only a handful of studies have investigated whether social factors (e.g., performing a task with another person) affect multisensory integration (Heed et al., 2010; Wahn et al., 2017a). This comes something as a surprise given that past work on joint action has demonstrated that social factors can have a significant impact on how individuals perceive an isolated visual event (Sebanz et al., 2003, 2006; Knoblich and Sebanz, 2006; Böckler et al., 2012; Karlinsky et al., 2017; Vesper et al., 2017). Also, people in everyday life routinely perform multisensory tasks with, or in the presence of, others. For instance, when eating a meal with a friend, visual, tactile, smell, and taste information are combined; and when attending a concert both visual (e.g., seeing the musicians) and auditory information (e.g., hearing the music) are processed in the presence of others.

To the best of our knowledge, the first study (Heed et al., 2010) that investigated the relationship between multisensory integration and joint task performance involved a tactile spatial localization task. Heed et al. (2010) required participants to indicate the location of a tactile stimulus while a visual stimulus was presented simultaneously in either the same (congruent) or different (incongruent) location. Past work has demonstrated that when participants perform this type of task alone, responses to the tactile stimulus are slower and less accurate if the visual stimulus appears at an incongruent location (Spence et al., 2004). Heed et al. (2010) found that this congruency effect was reduced when performing the task jointly, as the participant performing the tactile task "off-loaded" attending to the visual distractor to their partner. As a potential mechanism to explain this effect, Heed et al. (2010) suggested that the participant performing the tactile task co-represented (Sebanz et al., 2003) the partner's task and could hence better filter out the visual distractors from their own task representation. The process of co-representation (i.e., that co-actors take into account each other's tasks) has been proposed to occur automatically whenever co-actors perform tasks jointly (Sebanz et al., 2003) and to form the basis for more complex joint actions (Vesper et al., 2010). In a recent study, we replicated the finding by Heed et al. (2010) in a joint audiovisual congruency task (Wahn et al., 2017a). That is, we found that the negative effect of an incongruent visual stimulus on sound localization was reduced for participants performing the task jointly. Relatedly, Sellaro et al. (2018) found that such a division of labor of tasks also reduced interference in a purely visual picture–word interference task (for a recent review about the benefits of labor division in joint tasks, see Wahn et al., 2018).

While the above research has demonstrated that performing a task jointly does affect audiovisual and visuotactile integration in a spatial congruency task (Heed et al., 2010; Wahn et al., 2017a), it has not been investigated whether the effect of joint performance on multisensory integration can be generalized to other situations, particularly whether the results extend beyond the presentation of two solitary, static stimulus events. The stimuli one routinely encounters in everyday life are normally in motion because we are often in motion (e.g., walking, moving our head, and shifting our gaze several times a second) and the world around us is in motion, too (e.g., living animals move, water flows, and trees sway in the wind). An important extension of the previous work then is to test if multisensory integration with moving stimuli is affected by joint performance. That is, investigating this question would be informative as to whether multisensory integration of *moving* stimuli is also affected by joint task performance or whether the effect of joint task performance is specific to stationary spatial stimuli.

Soto-Faraco et al. (2002) introduced an audiovisual motion congruency task that is conceptually very similar but qualitatively distinct from the crossmodal congruency task with static stimuli (Soto-Faraco et al., 2002, 2004), but to date it has only been tested with individual, isolated participants. In the typical audiovisual motion task, a participant receives visual and auditory stimuli that either move together in the same direction (congruent presentations) or in opposite directions (incongruent presentations). The critical task is to judge the movement direction of the auditory stimuli. Results indicated that participants often failed to correctly identify the direction of sound motion when the direction of the visual motion was incongruent (e.g., leftward auditory motion and rightward visual motion). We viewed this paradigm as a logical next step to test whether joint task performance affects audiovisual integration using a task involving more ecologically valid stimuli (i.e., motion stimuli). That is, the audiovisual motion congruency task used by Soto-Faraco et al. (2002) represented only a minimal change (i.e., static stimuli are replaced by moving stimuli) relative to our earlier study using a spatial audiovisual congruency task (Wahn et al., 2017a).

Importantly, past work has also demonstrated that the effect of multisensory integration varies with the nature of the task. For instance, as discussed above, when judging the spatial direction of two auditory stimuli, irrelevant and incongruent visual stimuli have a negative effect on performance. Note, when the task is reversed, and one is required to determine the direction of two visual stimuli, incongruent auditory motion has no impact on performance. In general, the explanation for this asymmetry is that multisensory integration is preferentially biased toward the modality that provides the most reliable signal for the task at hand, in this case, spatial direction. In other words, vision provides a more reliable spatial signal than does audition, a point we are all too familiar with when we are trying to determine in a group whose phone is ringing; it is only when we see a person move that we localize the sound. Critically, this advantage of a visual signal over an auditory signal reverses when the task is to judge when, rather than where, two events have occurred. This was demonstrated by Morein-Zamir et al. (2003) who asked participants to judge which of two visual stimuli appeared first on a computer screen. They found that performance improved when an auditory click trailed the second visual stimulus, as if the second visual event was pulled toward the trailing auditory click. As the paradigm used by Soto-Faraco et al. (2002) can be readily adapted to that of Morein-Zamir et al. (2003), we examined if a joint task manipulation affects both dynamic spatial judgments and temporal judgments. That is, adapting the paradigm used by Soto-Faraco et al. (2002) to that of Morein-Zamir et al. (2003) only involves minimal changes (i.e., instead of judging the direction of two rapidly presented stimuli, participants are required to judge the temporal order of two stimuli), allowing for the specific targeting of the question of whether temporal judgments are affected by joint task performance in a within-subject design. In doing so, we can also assess whether potential effects of joint task performance are comparable both in situations where vision affects auditory judgments (i.e., sound direction) and when audition affects visual judgments (i.e., visual timing). Finally, it is worth noting that there is currently uncertainty in the literature as to whether joint task performance on multisensory integration affects how quickly people respond, their accuracy of response, or both. For instance, Heed et al. (2010) used inverse efficiency scores (i.e., a combined measure of response times and accuracy) to analyze their data, rendering the speed/accuracy question equivocal. Wahn et al. (2017a) analyzed response times and accuracy separately and found that joint task performance only affected response times. As crossmodal congruency effects on response times are often vulnerable to alternative explanations that do not demand an explanation in terms of multisensory integration (e.g., a race model explains why two congruent signals may result in faster responses time than either of them alone, see Miller, 1982; Stevenson et al., 2014) we aimed to test whether performing a task jointly affects perceptual judgments (i.e., perceptual accuracy) rather than response times. Crucially, in both tasks mentioned above (i.e., the motion discrimination task and temporal order judgment task), crossmodal effects were quantified via response accuracies.

To summarize, the current work aims to extend previous research on multisensory integration and joint task performance (Heed et al., 2010; Wahn et al., 2017a) in three ways: (1) through the use of moving rather than static stimuli in a crossmodal congruency task, (2) by investigating temporal crossmodal effects, and (3) by using tasks that quantify crossmodal effects with regard to response accuracy (rather than response time).

FIGURE 1 | Experimental setup. Response keys were marked by green stickers on the keyboards.

2. METHODS

2.1. Participants

Nineteen pairs of students (32 female and 6 male, $M = 19.58$ years, $SD = 1.44$ years) of the University of British Columbia participated in the present study. The participants provided their written informed consent to participate in this study and received course credits as compensation for their participation.

2.2. Experimental Setup

Pairs of participants were seated next to each other, 60 cm from a computer screen (resolution: 1920 x 1000 pixels, 64.13 x 33.40 visual degrees, 60 Hz refresh rate, model: ACER V243H, 24 inches) so that when they looked straight ahead they could see the left or right edge of the computer screen, respectively. The auditory stimuli were received via speakers (model: Dell A215) placed next to the computer screen. The speakers were positioned at a height so that the middle of the speakers would align with the middle of the screen and were about 80 cm apart from each other. In front of each participant, a QWERTY keyboard was positioned for making responses (see **Figure 1** for an overview of the experimental setup). The experiment was run on an Apple Mac Mini (2012 model), and we used its internal sound card to play the auditory stimuli.

2.3. Experimental Procedure

Each pair of participants performed the motion discrimination task and the temporal order judgment task. The order of tasks was counterbalanced across pairs so that half of the pairs started with the motion discrimination task and the remaining half began with the temporal order judgment task. In the following section we describe the procedure for each task separately. As a point of note, the participant performing the auditory motion discrimination task also performed the visual temporal order judgment task. The reasoning for this design choice was that, in both tasks, the crossmodal effects were expected to occur, and we planned to correlate the size of the crossmodal effects and the size of social effects across tasks.

2.3.1. Motion Discrimination Task

In the motion discrimination task, two beeps (duration: 50 ms) were presented, one from each speaker, one after the other in a rapid sequence (interstimulus interval: 100 ms) to create the apparent perception that stimuli were moving either in the left or right direction. The frequency of the two beeps was randomly selected out of a set of three frequencies (450, 500, and 550 Hz). Simultaneously with the presentation of the beeps, two flashes (duration: 50 ms) were presented that moved either in the same direction (congruent presentation) or opposite directions (incongruent presentation) of the auditory stream. The flashes (radius: 1.34 visual degrees) were presented at a distance of 15 visual degrees from a fixation dot (radius: 0.53 visual degrees) that was positioned in the center of the computer screen. In control trials, the flashes and beeps were presented asynchronously (for an overview of stimuli combinations see **Figure 2**). That is, in the asynchronous trials, the presentation of the flashes began 300 ms after the second beep was presented. For each trial, we

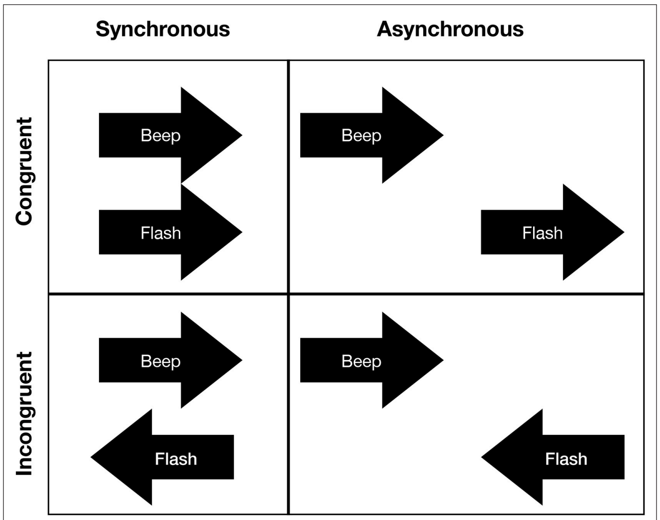


FIGURE 2 | Stimuli combinations for the motion discrimination task: Auditory and visual stimuli could either synchronously (upper left) or asynchronously (upper right) move in the same direction or opposite directions (lower left and lower right). If stimuli were presented asynchronously, the first flash was presented 300 ms after the second auditory stimulus.

randomly selected whether stimuli were presented synchronously or asynchronously.

One of the participants in the pair was required to perform the same auditory motion discrimination task as in the original study by Soto-Faraco et al. (2002). That is, they were required to indicate the movement direction of the beeps. For participants that sat on the left, they indicated the motion direction using the "A" key (for leftward motions) and the adjacent "S" (for rightward motions). For participants that sat on the right, they used the "K" (for leftward motions) and the adjacent "L" (for rightward motions). Responses were performed on the two keyboards placed in front of the participants (i.e., participants sitting on the left used the left keyboard, whereas participants sitting on the right used the right keyboard). While participants performed this task, they were also required to fixate the central dot on the computer screen. We did not explicitly instruct participants to turn their heads to the computer screen but often observed this to be the case as it is a more natural head position to fixate the central dot. To directly align their heads with the center of the screen, participants likely turned their heads by about 25 degrees. To ensure that participants maintained fixation, similar to Heed et al. (2010), there were a small number of catch trials (11 %), in which the central fixation dot would briefly flash (50 ms) and no other stimuli were presented. When this happened, the participant was required to press the "space" key if they sat on the left and the "enter" key if she/he sat on the right ("fixation control task"). Participants were instructed to prioritize accuracy over speed for their responses. As a point of note, any of the keys would end a trial regardless of the required task.

The other participant in the pair was required to perform the fixation control task and indicate the movement direction of the flashes ("visual motion discrimination task"). Again, depending on the seating position, either the "enter" or "space" key would be required for the fixation control task and either the "A" and "S" or the "K" and "L" keys would be required for indicating the moving direction of the visual stimuli.

Participants performed their assigned tasks either alone or jointly (see **Figure 3**). When they were alone in the room, they sat in the same seat that they occupied when performing the task jointly. The seating positions of participants performing the different tasks were counterbalanced across pairs.

As a point of note, as it has been done in earlier studies (Heed et al., 2010; Wahn et al., 2017a), in the data results section we only considered the response data of the participant performing the auditory motion discrimination task, as the crossmodal effects were expected to occur in the auditory motion discrimination task (i.e., the visual stimuli were expected to influence the auditory motion discrimination but not vice versa).

Testing involved two sets of three blocks: visual discrimination performed alone, auditory motion discrimination performed alone (by the other participant), and visual and auditory discrimination tasks performed simultaneously by the two participants together. The order of the conditions in a set was randomly selected and then repeated. Each block had 56 trials, composed of 8 fixation control trials and 48 motion discrimination task trials. Each block was composed of an equal number of trials for each combination of the factor levels of Synchrony (synchronous, asynchronous) and Congruency (congruent, incongruent) trials (e.g., 12 synchronous congruent trials and 12 synchronous incongruent trials). After the last required response on a trial, the program automatically continued to the next trial following a 1,000 ms break.

At the beginning of each block, the block type was announced on the screen ("Joint Block," "Individual Auditory Block," or "Individual Visual Block"), and participants were asked to contact the experimenter. The experimenter would then make the necessary setup adjustment (e.g., ask one of the participants to wait outside of the experimental room). The experimenter waited outside of the experimental room throughout testing.

The experiment was programmed using Python 2.7.3. It took about 20 min to complete.

2.3.2. Temporal Order Judgment Task

In the temporal order judgment task, two flashes (radius: 1.34 visual degrees, 5 ms) were presented in a rapid temporal sequence. The time between the flash presentations was randomly selected for each trial out of a set of four stimulus onset asynchronies (SOAs): 25, 50, 75, and 100 ms. One flash was presented below and one above the fixation dot (radius: 0.53 visual degrees) at a distance of 15 visual degrees. Whether the top or the bottom flash was presented first varied randomly between trials. Simultaneously with the first flash, a click sound (impulse tone, 5 ms) was presented as well. Depending on the type of trial, a second click was presented simultaneously with the second flash (baseline trial) or the second click trailed behind the second flash by 100 ms (trailing trial) (for an overview of all stimuli combinations, see **Figure 4**). One of the clicks was presented from the left loudspeaker and the other from the right loudspeaker. Baseline and trailing trials were selected randomly,

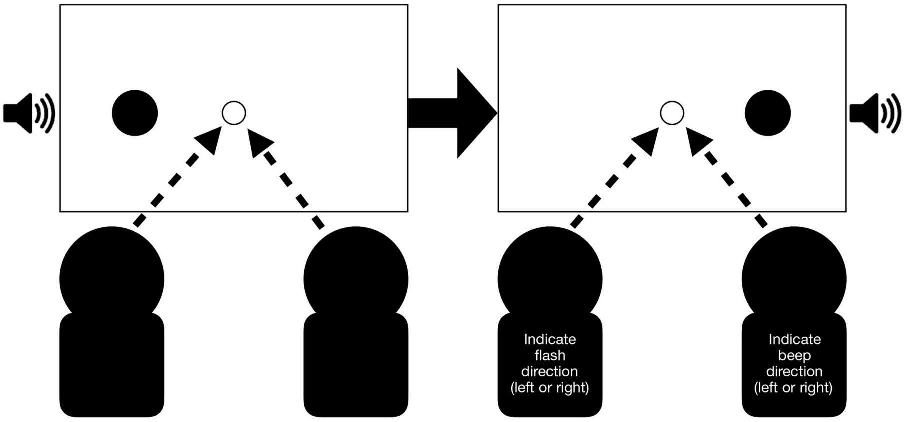


FIGURE 3 | Example of a synchronous congruent trial for the joint condition: Participants receive a flash and beep presented on the left side followed by a flash and beep presented on the right side. In this example pair, the left participant is required to indicate the flash direction (left or right) while the participant sitting on the right is required to indicate the beep direction (left or right). Arrows indicate that participants were also required to fixate the center of the screen. A trial is completed after both participants pressed a key.

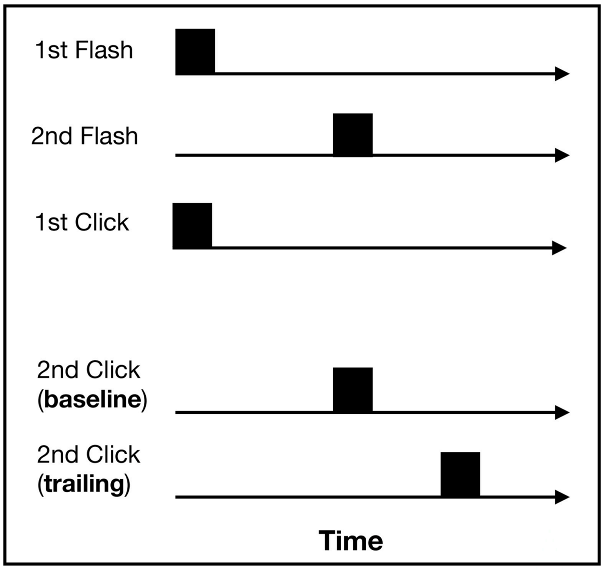


FIGURE 4 | Stimuli combinations for the temporal order judgment task: two auditory and two visual stimuli were presented. The first visual and auditory stimulus was presented at the same time. The second auditory stimulus was either presented together with the second flash (baseline trial) or 100 ms after the second flash (trailing trial).

as were which speaker delivered which sound. Given earlier findings by Morein-Zamir et al. (2003), we expected crossmodal effects to occur in the trailing trials as the trailing click should affect the temporal perception of the second flash, "pulling" the perception of the two flashes apart.

As a point of note, the stimuli presentations of the clicks (left and right) and flashes (top and bottom) were deliberately chosen to be orthogonal to avoid any spatial crossmodal influences between stimuli. Moreover, contrary to the stimuli presentations in the motion discrimination task above, the flashes remained on the screen after stimulus onset and only disappeared after the participants' responses. Also note that we deviated with regard to a few design choices from the original study by Morein-Zamir et al. (2003). The original study selected SOAs from a larger set with a wider range (12, 24, 36, 48, 72, 96, and 144 ms). We selected SOAs from a smaller range and set (25, 50, 75, and 100 ms) to reduce the overall number of trials. Moreover, the original study included several types of trailing trials (ranging from 75 to 600 ms). We only used one type of trailing trials, for which the crossmodal effect was the strongest in the original study (100 ms).

As for the motion discrimination task above, each participant in a pair was assigned to perform a task in one of the sensory modalities. In particular, one participant in the pair was required to indicate whether the upper or lower flash occurred first (visual temporal order judgment task), which was the same task as performed by participants in the original study (Morein-Zamir et al., 2003). If the participant was sitting on the left, she/he was required to use the "A" key to indicate that the top flash came first and the "Z" key to indicate that the bottom flash occurred first. If the participant was sitting on the right, she/he was required to use the "K" key to indicate that the top flash occurred first and the "M" key to indicate that the bottom flash occurred first. While performing the visual temporal order judgment task, the participant was also required to maintain fixation on the central fixation dot. As for the motion discrimination task above, during a small number of trials (3%), the participant was also required to perform the fixation control task. That is, the central fixation dot would briefly flash (50 ms) and no other stimuli were presented during such a trial. Also, for this task, we did not explicitly instruct participants to turn their heads toward the computer screen but often observed this to be the case as it is a more natural head position to fixate the central dot. For these fixation control task trials, depending on the seating position, the participant was required to press "space" (sitting of the left) or "enter" (sitting on the right).

The other participant in the pair was required to indicate which of the clicks occurred first (auditory temporal order judgment task). If the participant was sitting on the left, pressing the "A" key would be required to indicate that the click played on the left speaker occurred first and the adjacent "S" key when the click played on the right speaker occurred first. If the participant was sitting on the right, the "K" (for left) and "L" (for right) were required. While performing the auditory temporal order judgment task, the participant was also required to maintain fixation at the central dot and to also perform the fixation control task. As for the motion discrimination task above, all participants were instructed to prioritize accuracy over speed for their responses for all tasks.

As for the motion discrimination task above and in line with earlier studies investigating the effect of joint task performance on multisensory integration (Heed et al., 2010; Wahn et al., 2017a), we only considered the response data of participants performing the visual temporal order judgment task since crossmodal effects were only expected to occur in this task. Indeed, like Morein-Zamir et al. (2003), we did not collect trailing visual stimuli to assess the influence of visual signals on auditory temporal order judgments.

As before, the experiment was divided into two sets of three blocks: one block for each participant to perform the visual or auditory temporal order judgment task while alone in the room and one block for the two tasks to be performed simultaneously while together in the room (see **Figure 5**). Each block contained 136 trials. Four of these were trials for the fixation control task. Half of the remaining trials were trailing trials, and the other half were baseline trials. After responses were made, the program automatically continued to the next trial after a 1,000 ms break. Again, the experimenter waited outside of the experimental room throughout testing, making the necessary setup adjustment between blocks.

The experiment took about 40 min to complete. It was programmed in Python 2.7.3.

2.4. Data Pre-processing

For our data analysis later on, in line with earlier studies and as noted in the task procedure, we only considered data of the participants performing the tasks in which crossmodal effects were expected to occur (i.e., the auditory motion discrimination task and visual temporal order judgment task).

To briefly confirm this expectation, at least for the motion task (as an analysis for the auditory temporal order judgment task is not feasible as noted above), we assessed the performance in the visual motion discrimination task for the synchronous individual condition and found a high accuracy performance regardless of the type of presentations (congruent: $M = 0.92$ vs. incongruent: $M = 0.90$). We also found that there was no significant difference between congruent and incongruent presentations [$t_{(11)} = -1.16$, $p = 0.269$], suggesting that there were no crossmodal effects present in the visual motion discrimination task. In the following, only data from the auditory motion discrimination and visual temporal order judgment task were considered.

To assess participants' general performance accuracy for the two tasks we primarily considered for the analysis (i.e., the auditory motion discrimination task and visual temporal order judgment task), we used baseline data from the conditions where no crossmodal effects were expected to occur. For the auditory motion discrimination task in particular, we used the data from the incongruent asynchronous presentations when a participant performed their task alone in the room. For the visual temporal order judgment task, we used the baseline trials with the longest SOA (100 ms) when a participant was alone in the room. For both these situations, and for the fixation control tasks, we set the inclusion criteria to a performance above 70%.

We aimed to match the sample size of our current study to the sample size of earlier studies investigating social manipulations in crossmodal tasks, which was 11 in the case of Heed et al. (2010) and 12 in the case of Wahn et al. (2017a). Moreover, we counterbalanced the seating position and task order across pairs such that we have an equal number of pairs for each combination of these factors. We also sought to have a sample of participants that were able to accurately perform the motion discrimination task, temporal order judgment task, and fixation control task. Our data collection ran until all these criteria were fulfilled for a sample size of 12 pairs (21 females and 3 males, $M = 19.67$ years, $SD = 1.68$ years). An additional 7 pairs (11 females, 3 males, $M = 19.42$ years, $SD = 1.03$ years) did not fulfill our inclusion criteria.

The fixation control task was performed at a high accuracy in the final sample both in the motion discrimination experiment ($M = 97\%$ correct, $SD = 4.81\%$) and temporal order judgment experiment ($M = 96\%$ correct, $SD = 3.08\%$).

3. RESULTS

3.1. Auditory Motion Discrimination Task

Based on Soto-Faraco et al. (2002), we expected the factors Congruency and Synchrony to interact. The rationale is that the performance difference between incongruent and congruent presentations should be larger for the synchronous than the asynchronous presentations because it is during the synchronous presentations that the incongruent visual signals should distract participants from accurately judging the sound motion. For social factors to influence this crossmodal effect, we would expect a significant interaction between the factors Congruency, Synchrony, and Social Condition.

Figure 6 (upper panels) displays the response accuracy for all combinations of these factors, including the task order. On a descriptive level, we observed large differences between congruent and incongruent presentations for the synchronous trials, suggesting that the crossmodal effects that found in the original study Soto-Faraco et al. (2002) were replicated in the present study. With regard to our social manipulation, we observed that the difference between congruent and incongruent presentations for synchronous trials was not modulated by whether a task was performed alone or jointly. Lastly, we did not observe any order effects (i.e., that the pattern of results for performing the motion discrimination task first or second did not differ).

To confirm whether these observations were significant, we analyzed participants' performance using a four-factorial ANOVA with the response accuracy as the dependent variable.

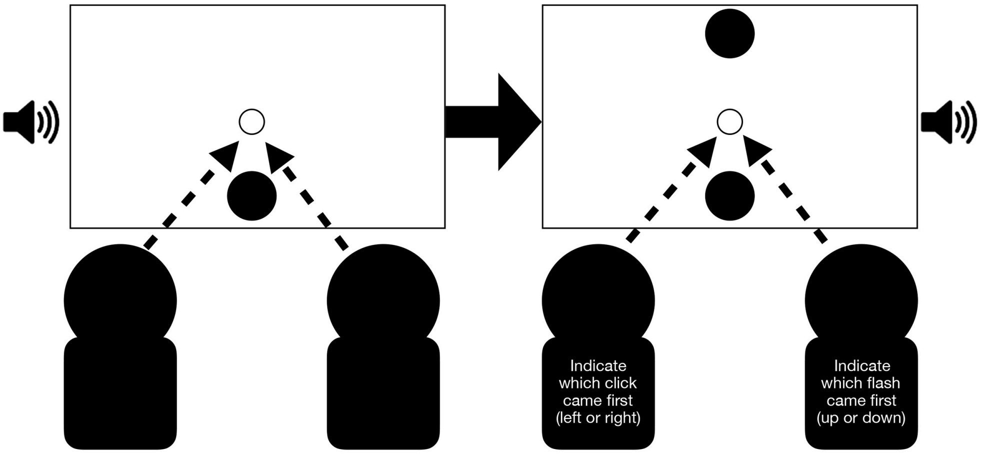


FIGURE 5 | Example of baseline trial in the joint condition: Participants first receive a flash presented at the bottom and simultaneously a click sound on the left. Then, a second flash is presented (the first one remains on the computer screen) simultaneously with a second click on the right. In this example seating arrangement, the left participant is required to indicate which click came first and the right participants is required to indicate which flash came first. Arrows indicate that participants were also required to fixate the center of the screen. A trial is completed after both participants pressed a key.

The three within-subject factors were Congruency (Incongruent and Congruent), Synchrony (Synchrony and Asynchrony), and Social Condition (Alone and Joint). The between-subject factor was the Task Order (First and Second). We found main effects for the factors Synchrony [$F_{(1,\,10)} = 41.60$, $p < 0.001$, $\eta^2_G = 0.43$] and Congruency [$F_{(1,\,10)} = 65.55$, $p < 0.001$, $\eta^2_G = 0.63$]. Replicating crossmodal effects from the original study (Soto-Faraco et al., 2002), we found an interaction effect between these two factors [$F_{(1,\,10)} = 184.27$, $p < 0.001$, $\eta^2_G = 0.58$]. Given that our ANOVA only involved factors with two levels, there were no follow-up tests required since the interaction effect already tested the pairwise comparison. To describe this interaction in more detail, the difference in performance accuracies between congruent and incongruent presentations was significantly larger for synchronous presentations than for asynchronous presentations. All other effects in the ANOVA were not significant (all $ps > 0.129$).

As the absence of a significant interaction effect between the factors Congruency, Synchrony, and Condition [$F_{(1,\,10)} = 0.11$, $p = 0.746$] suggested that performing a task jointly did not affect audiovisual integration, we also computed a Bayes factor using the R package "BayesFactor" (Morey et al., 2015) for this effect to assess how much more likely the null hypothesis was relative to the alternative hypothesis. We found a Bayes factor of 0.30, meaning that our data was 1/0.30 or 3.33 times more likely under the null hypothesis than the alternative hypothesis.

To test for the possibility of whether an auditory stimulus moving toward where the participant was sitting or away from the participant may have interacted with our Social Condition factor, we repeated the above ANOVA with the additional factor Auditory Moving Direction (Away and Toward). Apart from the effects already found above (i.e., a significant main effects for the factors Congruency [$F_{(1,\,10)} = 65.55$, $p < 0.001$, $\eta^2_G = 0.54$] and Asynchrony [$F_{(1,\,10)} = 41.60$, $p < 0.001$, $\eta^2_G = 0.34$] and a significant interaction effect between these two effects [$F_{(1,\,10)} = 184.27$, $p < 0.001$, $\eta^2_G = 0.49$], no other effects in the ANOVA were significant (all $ps > 0.068$).

Apart from assessing response accuracy, we also repeated the same ANOVA above for the response times as the dependent variable. We found no significant effects for this ANOVA. In particular with regard to crossmodal effects, there was no significant interaction effect between the factors Congruency and Synchrony [$F_{(1,\,10)} = 1.18$, $p = 0.302$], suggesting there were no crossmodal effects present from the perspective of the response times.

In sum, we replicated earlier crossmodal effects by Soto-Faraco et al. (2002), finding that for synchronous presentations visual stimuli affected auditory motion judgments more than for asynchronous presentations. We did not find that performing a task jointly modulated this effect. In fact, our calculated Bayes factor suggests that the null hypothesis that there is no effect is considerably more likely than the alternative hypothesis.

3.2. Visual Temporal Order Judgment Task

For analyzing the data of the visual temporal order judgment task, we followed the same analysis procedure as in the original study by Morein-Zamir et al. (2003). That is, using a logistic regression, we fitted psychometric curves to each of the participants' responses, separately for each condition. Based on these fits, we extracted for each participant the just noticeable difference (JND). These JNDs were used as a dependent variable for our further analyses.

To replicate the crossmodal effect found by Morein-Zamir et al. (2003), we expected a main effect for the factor Timing with the levels baseline and trailing. That is, we expected that participants were significantly better at judging which of the two flashes occurred first for trailing trials compared to baseline trials. For a social effect to occur, we expected an interaction effect between the factors Social Condition (Alone, Joint) and Timing (Baseline, Trailing). Plotting the averaged JNDs as a

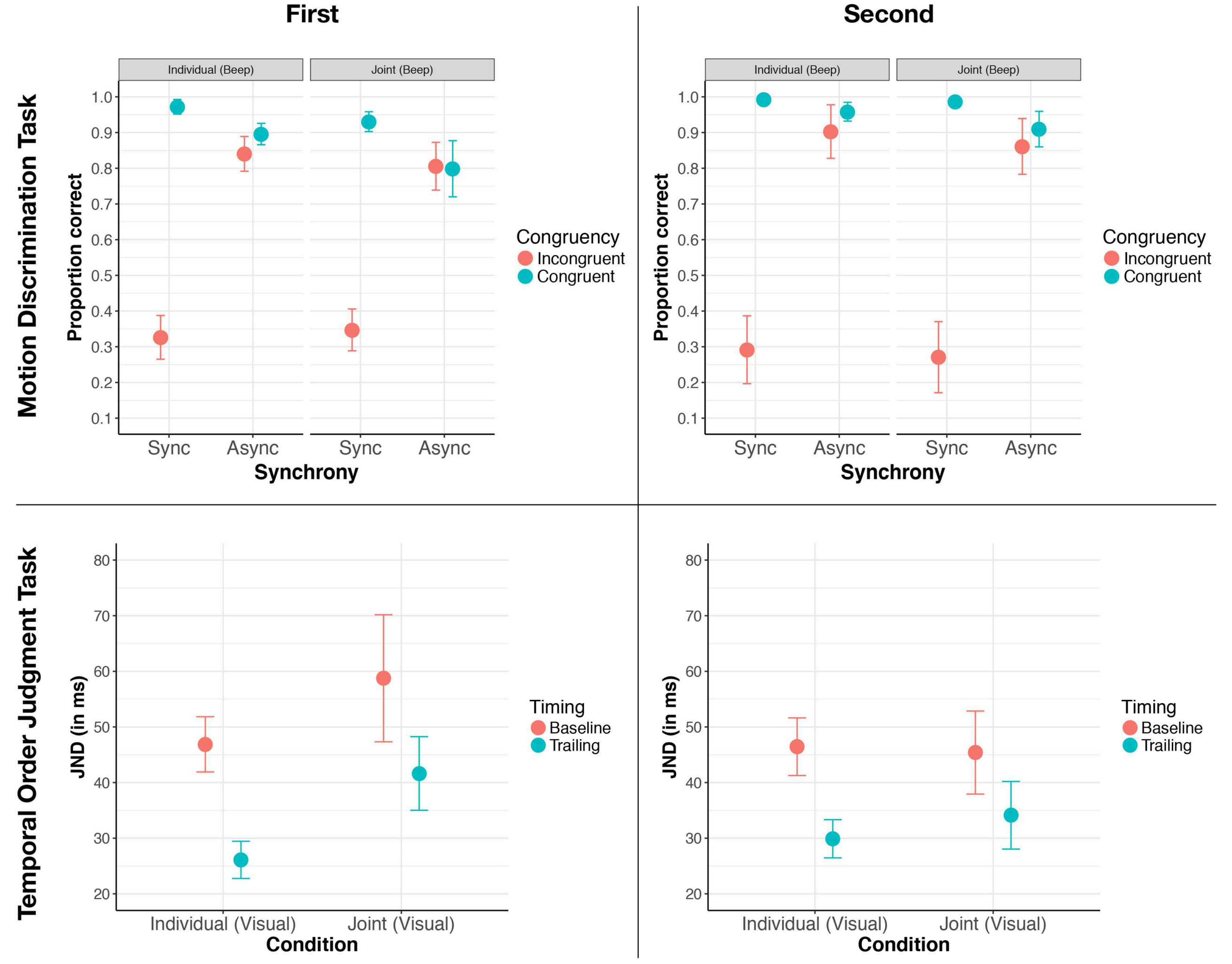


FIGURE 6 | Results overview for the motion discrimination task (upper two panels) and the temporal order judgment task (lower two panels) when the respective task is performed first (left column) or second (right column). Error bars are standard error of the mean in all panels.

function of these factors, including Task Order (see **Figure 6**, lower panels), we observed large differences between baseline and trailing trials, suggesting that we replicated the earlier crossmodal effect. Yet, the size of this crossmodal effect (i.e., the difference in JND between baseline and trailing trials) did not appear to be modulated by whether the task was performed jointly or alone. However, there was the suggestion that the effect of the Social Condition was modulated by Task Order, with the JNDs somewhat raised for the joint condition compared to the individual condition when the temporal order judgment task was performed first.

To assess whether these observations were statistically significant, we performed a three factorial ANOVA with the JNDs as the dependent variable and the within-subject factors Social Condition (Alone, Joint) and Timing (Baseline, Trailing) and the between-subject factor Task Order. We found a main effect of Timing [$F_{(1,\,10)} = 39.05$, $p < 0.001$, $\eta^2_G = 0.24$]. All the other effects were not significant (all $ps > 0.124$). There were no interaction effects, including, most importantly, no significant Social Condition x Timing interaction effect, [$F_{(1,\,10)} = 1.06$, $p = 0.327$]. For this interaction, we again computed a Bayes factor to assess how more likely the null hypothesis is compared to the alternative hypothesis given the present data. We observed a Bayes factor of 0.47, meaning that our data were 1/0.47 or 2.15 times more likely under the null hypothesis than the alternative hypothesis.

In sum, we replicated the crossmodal effect found by Morein-Zamir et al. (2003). As for the motion discrimination task above, we found that audiovisual integration was not affected by performing the task jointly rather than alone. In fact, the null hypothesis that there is no effect was more than two times more likely than the alternative hypothesis.

Similar to the auditory motion discrimination task, we also tested whether the click starting position (either starting on the participant's side or opposite side) interacted with our social manipulation. For this purpose, we repeated the ANOVA above with the additional factor Auditory Starting Position (Same and Opposite). We again found a significant main effect of Timing [$F_{(1,\,10)} = 67.02$, $p < 0.001$, $\eta^2_G = 0.26$]. Moreover, we found a significant main effect of Auditory Starting Position [$F_{(1,\,10)} = 7.70$, $p = 0.012$, $\eta^2_G = 0.03$] and a significant interaction effect between Timing and Auditory Starting Position [$F_{(1,\,10)} = 7.84$, $p = 0.019$, $\eta^2_G = 0.03$]. Yet, none of the effects involving the factor Social Condition (Alone and Joint) were significant (all $ps > 0.155$).

Given that we replicated earlier crossmodal effects in both tasks, we also correlated the sizes of these effects across tasks. For each task we averaged the data across the levels of all factors except for Timing in the temporal order judgment task and Congruency in the motion discrimination task. For the temporal order judgment task, we then computed the difference between the baseline and trailing condition. For the motion discrimination task, we computed the difference between the congruent and incongruent condition. Correlating these differences, we found a moderately sized correlation, which was not significant [$r = -0.39$, $t_{(10)} = -1.35$, $p = 0.204$]. For this correlation, we found a Bayes factor of 0.80, meaning that our data were 1/0.80 or 1.15 times more likely under the null hypothesis than the alternative hypothesis.

4. DISCUSSION

In the present study, we replicated earlier crossmodal effects in a motion discrimination task (Soto-Faraco et al., 2002) and temporal order judgment task (Morein-Zamir et al., 2003); auditory judgments of motion were affected by visual input, and visual judgments of timing were affected by auditory input. In the case of the motion discrimination task, we found that participants' performances were significantly worse when incongruent rather than congruent visual information was presented. For the temporal order judgment task, participants were better at judging the order of the flashes if a click trailed the second flash. These findings demonstrate that these known crossmodal effects were robust, persisting despite the many design changes we made, the most profound ones being that we introduced social situations where two participants performed their respective tasks together.

On this last score, despite the introduction of joint task situations, and contrary to previous findings that a joint performance modulated crossmodal spatial congruency effects (Heed et al., 2010; Wahn et al., 2017a), we found that our robust crossmodal motion and timing effects were not modulated by joint task performance. These findings highlight the importance of investigating the effects of social factors across a range of crossmodal paradigms. In other words, as we had noted in the introduction, one cannot assume that because social factors impact crossmodal performance in one particular paradigm, social factors will affect crossmodal performance in all situations. Below, we speculate why the present tasks are resistant to social manipulations and how one might test our proposals in the future.

A major difference between the present task and crossmodal congruency tasks investigated earlier is that crossmodal effects in the present study were quantified in both tasks via perceptual judgments (i.e., response accuracies), while in the earlier studies effects were quantified (at least in part) with response times. Moreover, in the audiovisual crossmodal congruency task investigated earlier (Wahn et al., 2017a), the effect of task co-performance was also only present for the response times while response accuracies were unaffected. Given that response accuracies were also not affected by joint task performance in the present study for both tasks, one could also suggest that performing a task jointly specifically affected crossmodal effects quantified with response times while performance accuracies remain unaffected. An outstanding question for future investigation is whether those past response time crossmodal effects reflect multisensory integration at all or merely the speed at which one of the signals reaches the response threshold. The fact that no social effect has been observed in response accuracy suggests that there may not be any social effect on multisensory integration.

An alternative explanation for the divergent findings between the present data and previous work may rest with the difference in task demands between the present study and earlier studies. In particular, in earlier studies participants were required to localize static stimuli whereas in the present study participants were required to judge the movement direction of stimuli and their temporal order. Given this difference between tasks, one possibility could be that participants are only able to "off-load" stimuli to a co-actor if the task primarily involves static spatial stimuli (as in the earlier studies Heed et al., 2010; Wahn et al., 2017a) while this is not possible for moving stimuli or temporal stimuli. In other words, stimuli may be required to be spatial and static for co-actors to be able to "off-load" these stimuli to other co-actors. Possibly, the mechanism of task co-representation, which was suggested to have allowed participants to filter out distracting stimuli in earlier studies; (Heed et al., 2010; Wahn et al., 2017a) could be specific to static spatial stimuli. Future studies could test this proposal by investigating whether joint task performance also does not affect multisensory integration in other tasks requiring spatial processing of moving stimuli (e.g., in an audiovisual bounce-inducing effect; Sekuler, 1997; Grassi and Casco, 2009) or temporal processing (e.g., in the sound-induced flash illusion; Shams et al., 2002).

Related to the difference in task demands between the present study and those earlier, whether joint task performance affects multisensory integration or not may also depend on the strength of the integration of the investigated multisensory effect. In particular, for the motion discrimination task, the integrated moving stimuli may more strongly be integrated (and hence less susceptible to effects of joint task performance) as the received stimuli provide more cues (i.e., spatial *and* motion information) to be integrated. Similarly, for the temporal order judgment task, the mere fact of presenting two audiovisual stimuli may have resulted in a stronger integration of stimuli that is less susceptible to the effects of joint task performance. Future studies could test this proposal by investigating whether joint performance task performance also does not affect multisensory integration in other tasks that involve more richer audiovisual stimuli (than static spatial stimuli). Another difference to consider, which is specific to the temporal order judgment task in relation to the spatial congruency tasks in earlier studies (Heed et al., 2010; Wahn et al., 2017a), concerns the direction of the crossmodal effects (i.e., whether stimuli in the visual sensory modality influence processing stimuli in a different sensory modality or vice versa). That is, in the temporal order judgment task crossmodal effects were present for stimuli in the visual sensory modality (i.e., auditory information influenced visual processing) whereas in the earlier studies crossmodal effects were either present in the auditory or tactile sensory modalities due to an influence of visual stimuli. Hence, one could raise the possibility that joint task performance only affects multisensory integration in tasks, where the visual sensory modality is affecting processing in other sensory modalities but not vice versa. Yet, given that we did not find an effect of joint task performance for the audiovisual motion discrimination task—a task in which visual stimuli affect auditory processing—this proposal may only apply to static stimuli in crossmodal tasks.

In summary, we successfully replicated earlier crossmodal temporal and motion effects, in which participants were required to perform perceptual judgments. Yet, the present work fails to find evidence that joint task performance modulates these replicated crossmodal effects. Given that earlier studies found an effect of joint task performance for crossmodal spatial congruency tasks (Heed et al., 2010; Wahn et al., 2017a), we suggest that the effect of joint task performance on crossmodal tasks could potentially depend on how crossmodal effects are quantified (i.e., via responses times or accuracies) and task demands (i.e., whether tasks require processing stimuli in terms of location, motion, or timing).

DATA AVAILABILITY STATEMENT

The datasets collected for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of British Columbia's ethics committee. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

BW, JD, and AK: Study design, wrote the manuscript and revised the manuscript. BW: Programmed the experiments and analyzed the data.

FUNDING

We acknowledge the support of a DFG research fellowship (WA 4153/2-1) awarded to BW, and a NSERC Discovery grant and SSHRC Insight grant awarded to AK.

ACKNOWLEDGMENTS

We want to thank Michelle Tomaszewski, Jane Kim, and Katherine Chow with their help with the data collection.

REFERENCES

Alsius, A., Navarra, J., Campbell, R., and Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demands. *Curr. Biol.* 15, 839–843. doi: 10.1016/j.cub.2005.03.046

Bertelson, P., Vroomen, J., De Gelder, B., and Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Percept. Psychophys.* 62, 321–332. doi: 10.3758/BF03205552

Böckler, A., Knoblich, G., and Sebanz, N. (2012). Effects of a coactor's focus of attention on task performance. *J. Exp. Psychol. Hum. Percept. Perform.* 38:1404. doi: 10.1037/a0027523

Grassi, M., and Casco, C. (2009). Audiovisual bounce-inducing effect: Attention alone does not explain why the discs are bouncing. *J. Exp. Psychol. Hum. Percept. Perform.* 35:235. doi: 10.1037/a0013031

Guski, R., and Troje, N. F. (2003). Audiovisual phenomenal causality. *Percept. Psychophys.* 65, 789–800. doi: 10.3758/BF03194815

Heed, T., Habets, B., Sebanz, N., and Knoblich, G. (2010). Others' actions reduce crossmodal integration in peripersonal space. *Curr. Biol.* 20, 1345–1349. doi: 10.1016/j.cub.2010.05.068

Holmes, N. P., and Spence, C. (2005). Multisensory integration: space, time and superadditivity. *Curr. Biol.* 15, R762–R764. doi: 10.1016/j.cub.2005.08.058

Karlinsky, A., Lam, M. Y., Chua, R., and Hodges, N. J. (2017). Whose turn is it anyway? the moderating role of response-execution certainty on the joint simon effect. *Psychol. Res.* 83, 833–841. doi: 10.1007/s00426-017-0901-7

Knoblich, G., and Sebanz, N. (2006). The social nature of perception and action. *Curr. Dir. Psychol. Sci.* 15, 99–104. doi: 10.1111/j.0963-7214.2006.00415.x

Meredith, M. A., Nemitz, J. W., and Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J. Neurosci.* 7, 3215–3229. doi: 10.1523/JNEUROSCI.07-10-03215.1987

Meredith, M. A. and Stein, B. E. (1996). Spatial determinants of multisensory integration in cat superior colliculus neurons. *J. Neurophysiol.* 75, 1843–1857. doi: 10.1152/jn.1996.75.5.1843

Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals. *Cogn. Psychol.* 14, 247–279. doi: 10.1016/0010-0285(82)90010-X

Morein-Zamir, S., Soto-Faraco, S., and Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Cogn. Brain Res.* 17, 154–163. doi: 10.1016/S0926-6410(03)00089-2

Morey, R. D., Rouder, J. N., and Jamil, T. (2015). *Bayesfactor: Computation of Bayes Factors for Common Designs*. R package version 0.9. 12-2 [Computer software]. Available online at: https://CRAN.R-project.org/package=BayesFactor

Santangelo, V., and Macaluso, E. (2012). "spatial attention and audiovisual processing," in *The New Handbook of Multisensory Processes*, ed B. E. Stein (Cambridge, MA: The MIT Press), 359–370.

Santangelo, V., and Spence, C. (2008). Crossmodal attentional capture in an unspeeded simultaneity judgement task. *Vis. Cogn.* 16, 155–165. doi: 10.1080/13506280701453540

Santangelo, V., and Spence, C. (2009). Crossmodal exogenous orienting improves the accuracy of temporal order judgments. *Exp. Brain Res.* 194, 577–586. doi: 10.1007/s00221-009-1734-4

Sebanz, N., Bekkering, H., and Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends Cogn. Sci.* 10, 70–76. doi: 10.1016/j.tics.2005.12.009

Sebanz, N., Knoblich, G., and Prinz, W. (2003). Representing others' actions: just like one's own? *Cognition* 88, B11–B21. doi: 10.1016/S0010-0277(03)00043-X

Sekuler, R. (1997). Sound alters visual motion perception. *Nature* 385:308. doi: 10.1038/385308a0

Sellaro, R., Treccani, B., and Cubelli, R. (2018). When task sharing reduces interference: evidence for division-of-labour in stroop-like tasks. *Psychol. Res.* doi: 10.1007/s00426-018-1044-1. [Epub ahead of print].

Shams, L., Kamitani, Y., and Shimojo, S. (2002). Visual illusion induced by sound. *Cogn. Brain Res.* 14, 147–152. doi: 10.1016/S0926-6410(02)00069-1

Soto-Faraco, S., Lyons, J., Gazzaniga, M., Spence, C., and Kingstone, A. (2002). The ventriloquist in motion: illusory capture of dynamic information across sensory modalities. *Cogn. Brain Res.* 14, 139–146. doi: 10.1016/S0926-6410(02)00068-X

Soto-Faraco, S., Spence, C., Lloyd, D., and Kingstone, A. (2004). Moving multisensory research along: motion perception across sensory modalities. *Curr. Dir. Psychol. Sci.* 13, 29–32. doi: 10.1111/j.0963-7214.2004.01301008.x

Spence, C., Pavani, F., and Driver, J. (2004). Spatial constraints on visual-tactile cross-modal distractor congruency effects. *Cogn. Affect. Behav. Neurosci.* 4, 148–169. doi: 10.3758/CABN.4.2.148

Stevenson, R. A., Ghose, D., Fister, J. K., Sarko, D. K., Altieri, N. A., Nidiffer, A. R., et al. (2014). Identifying and quantifying multisensory integration: a tutorial review. *Brain Topogr.* 27, 707–730. doi: 10.1007/s10548-014-0365-7

Vercillo, T., and Gori, M. (2015). Attention to sound improves auditory reliability in audio-tactile spatial optimal integration. *Front. Integr. Neurosci.* 9:34. doi: 10.3389/fnint.2015.00034

Vesper, C., Abramova, E., Bütepage, J., Ciardo, F., Crossey, B., Effenberg, A., et al. (2017). Joint action: Mental representations, shared information and general mechanisms for coordinating with others. *Front. Psychol.* 7:2039. doi: 10.3389/fpsyg.2016.02039

Vesper, C., Butterfill, S., Knoblich, G., and Sebanz, N. (2010). A minimal architecture for joint action. *Neural Netw.* 23, 998–1003. doi: 10.1016/j.neunet.2010.06.002

Vroomen, J., Bertelson, P., and De Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Percept. Psychophys.* 63, 651–659. doi: 10.3758/BF03194427

Wahn, B., Keshava, A., Sinnett, S., Kingstone, A., and König, P. (2017a). "Audiovisual integration is affected by performing a task jointly," in *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (Austin, TX), 1296–1301.

Wahn, B., Kingstone, A., and König, P. (2018). Group benefits in joint perceptual tasks–a review. *Ann. N. Y. Acad. Sci.* 1426, 166–178. doi: 10.1111/nyas.13843

Wahn, B., and König, P. (2015a). Audition and vision share spatial attentional resources, yet attentional load does not disrupt audiovisual integration. *Front. Psychol.* 6:1084. doi: 10.3389/fpsyg.2015.01084

Wahn, B., and König, P. (2015b). Vision and haptics share spatial attentional resources and visuotactile integration is not affected by high attentional load. *Multisensory Res.* 28, 371–392. doi: 10.1163/22134808-00002482

Wahn, B., and König, P. (2016). Attentional resource allocation in visuotactile processing depends on the task, but optimal visuotactile integration does not depend on attentional resources. *Front. Integr. Neurosci.* 10:13. doi: 10.3389/fnint.2016.00013

Wahn, B., Murali, S., Sinnett, S., and König, P. (2017b). Auditory stimulus detection partially depends on visuospatial attentional resources. *Iperception* 8:2041669516688026. doi: 10.1177/2041669516688026

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Human Neuroscience

REVIEW
published: 28 February 2020
doi: 10.3389/fnhum.2020.00039

Hyperscanning: A Valid Method to Study Neural Inter-brain Underpinnings of Social Interaction

Artur Czeszumski[1]*, Sara Eustergerling[1], Anne Lang[1], David Menrath[1], Michael Gerstenberger[1], Susanne Schuberth[1], Felix Schreiber[1], Zadkiel Zuluaga Rendon[1] and Peter König[1,2]

[1] *Institute of Cognitive Science, Universität Osnabrück, Osnabrück, Germany,* [2] *Institut für Neurophysiologie und Pathophysiologie, Universitätsklinikum Hamburg-Eppendorf, Hamburg, Germany*

Social interactions are a crucial part of human life. Understanding the neural underpinnings of social interactions is a challenging task that the hyperscanning method has been trying to tackle over the last two decades. Here, we review the existing literature and evaluate the current state of the hyperscanning method. We review the type of methods (fMRI, M/EEG, and fNIRS) that are used to measure brain activity from more than one participant simultaneously and weigh their pros and cons for hyperscanning. Further, we discuss different types of analyses that are used to estimate brain networks and synchronization. Lastly, we present results of hyperscanning studies in the context of different cognitive functions and their relations to social interactions. All in all, we aim to comprehensively present methods, analyses, and results from the last 20 years of hyperscanning research.

Keywords: hyperscanning, social cognition, joint action, EEG, MEG, fMRI, fNIRS, social interactions

Edited by:
Zhen Yuan,
University of Macau, China

Reviewed by:
Joy Hirsch,
Yale University, United States
Tong Boon Tang,
University of Technology Petronas,
Malaysia

***Correspondence:**
Artur Czeszumski
aczeszumski@uos.de

Specialty section:
This article was submitted to Brain Imaging and Stimulation, a section of the journal Frontiers in Human Neuroscience

Received: *13 November 2019*
Accepted: *27 January 2020*
Published: *28 February 2020*

Citation:
Czeszumski A, Eustergerling S, Lang A, Menrath D, Gerstenberger M, Schuberth S, Schreiber F, Rendon ZZ and König P (2020) Hyperscanning: A Valid Method to Study Neural Inter-brain Underpinnings of Social Interaction. Front. Hum. Neurosci. 14:39. doi: 10.3389/fnhum.2020.00039

INTRODUCTION

The importance of social interaction for the development and maintenance of the human self was already highlighted in Greek philosophy and has been discussed ever since. Nevertheless, the field of cognitive neuroscience has only started to investigate brain activity during social interaction in the last decades. Typically, only the brain of one of the involved participants and thus only one part of the dyadic or group interaction was recorded at a time. The insight such experiments may provide is therefore limited. To examine social interactions as a whole, the idea of hyperscanning, i.e., measuring the activity of multiple brains simultaneously, has originated. The significant advantage of this technique is that it allows the investigation of real-time dynamics between two or more interacting brains (Hari and Kujala, 2009; Hari et al., 2013). In contrast to classic experimental paradigms that measure the brain activity of single participants during social interaction, simultaneously measuring the brain activity of several interacting participants allows for the investigation of intra- and inter-brain neural relations (Schilbach et al., 2013). The hyperscanning techniques thus offer a new approach to account for the complexity of joint action, i.e., its spontaneity, reciprocity, and multimodality, which constitutes a big challenge for its neuroscientific examination.

In the current paper, we have reviewed existing literature and evaluated the current state of the hyperscanning method. We performed extensive literature research to identify the most critical peer-reviewed studies that used hyperscanning as a method to investigate human social cognition.

In our review, we have had two primary goals. First, we reviewed the methods and types of analysis that are used in the hyperscanning field. Second, we reviewed cognitive functions and their neural underpinnings that are investigated with the hyperscanning method.

TYPE OF METHODS

In the last century, a large variety of methods to measure brain activity have been developed. The most popular ways to measure brain activity used in the cognitive neuroscience field are Electroencephalography (EEG) (Luck and Hillyard, 1994), Magnetoencephalography (MEG) (Baillet, 2017), Functional magnetic resonance imaging (fMRI) (Eisenberger, 2003), and functional near-infrared spectroscopy (fNIRS) (Ferrari and Quaresima, 2012). Each of these have their advantages and disadvantages, which can help us to further understand different brain functions. Primarily, when focused on Hyperscanning, their specific assets, like temporal and spatial resolution as well as mobility, are of value. We have reviewed here all of these in the context of Hyperscanning research.

fMRI

Functional magnetic resonance imaging (fMRI) is a method that indirectly measures neural brain activity. Namely, it measures it by detecting changes associated with blood flow, which is the blood-oxygen-level-dependent (BOLD) contrast (Glover, 2011). Since the last decade of the twentieth century, fMRI has become one of the most popular methods used in cognitive neuroscience. Its most important advantage is the spatial resolution. Standard fMRI scanners estimate brain activity with a spatial resolution of 3 mm, and specialized instruments push the limits toward the sub-millimeter range.

In comparison to other methods discussed below, it is the best method to determine where in the brain something has happened. Additionally, it is the non-invasive method of choice for measuring deep brain structures. However, because it uses blood flow to estimate neural activity, its temporal resolution does not compare to M/EEG (Glover, 2011). Moreover, to measure the BOLD signal, participants are required to stay stable in a laying position within a scanner (**Figure 1A**). This low mobility of the experimental tools makes it not suitable for investigating social interactions in naturalistic and ecologically valid setups. Despite low mobility, the first-ever hyperscanning study was an fMRI study. Montague (2002) performed a successful feasibility study to link participants in two scanners. To tackle the problem that occurs when two scanners are required to complete a study, King-Casas (2005) conducted a study using scanners in Texas and California linked via the Internet. Afterward, other studies were performed in facilities that possessed two scanners; however, until now, only a few studies have tried to investigate social interaction with the fMRI hyperscanning method (Tomlin, 2006; Saito et al., 2010; Schippers et al., 2010; Tanabe et al., 2012; Tomlin et al., 2013; Spiegelhalder et al., 2014; Koike et al., 2016, 2019; Abe et al., 2019). One reason for that might be difficulty

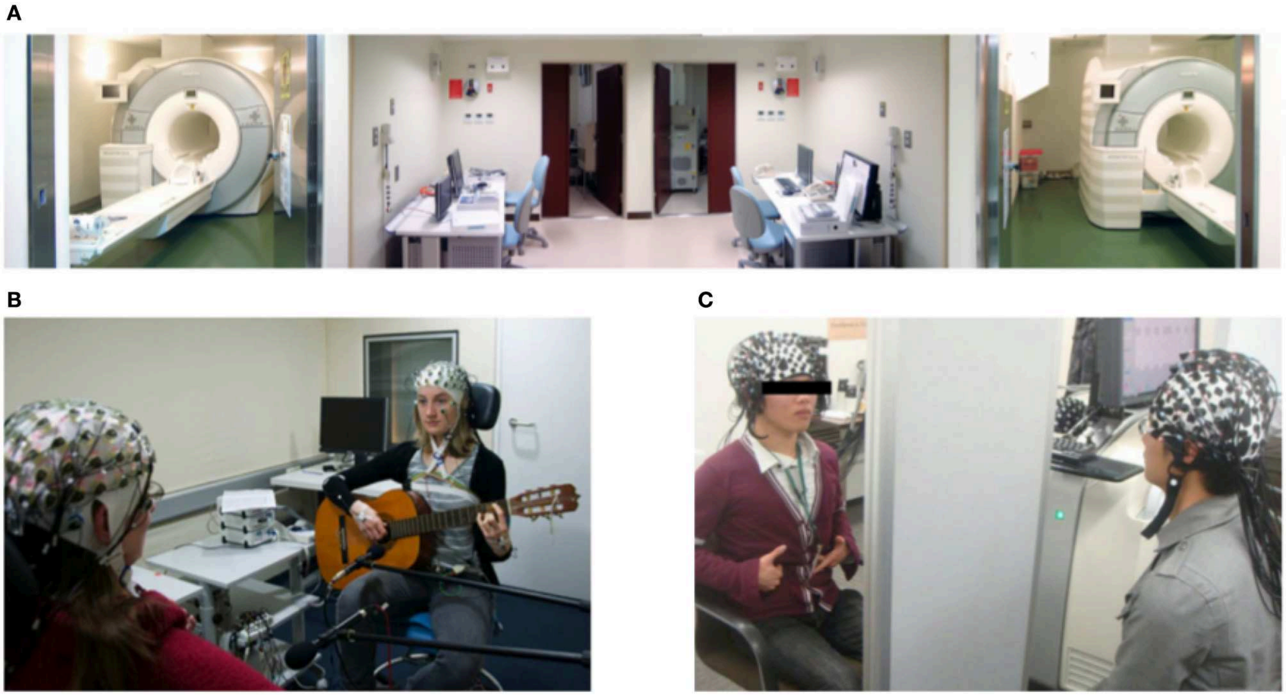


FIGURE 1 | Neuroimaging methods used in Hyperscanning. **(A)** From Koike et al. (2019). View of the dual fMRI facility used to study mutual gaze. **(B)** From Acquadro et al. (2016). EEG measurement of two guitar players. **(C)** From Osaka et al. (2015). fNIRS set up used to study cooperative singing. All parts reproduced/adapted under CC licenses.

in creating experimental paradigms that involve interaction between participants without movement and communication. Another reason might be that the complexity of fMRI data requires the development of new types of analysis that are suitable to answer questions about between-brain relations. It is sensible to say that the value of each of these studies is excellent, and more studies are required because fMRI has a good spatial resolution. Furthermore, this method could be of great value if combined with EEG to surmount poor temporal resolution (Koike et al., 2015).

EEG/MEG

One of the oldest methods to measure activity in the brain is electroencephalography (EEG). In comparison to fMRI and fNIRS, it measures neural activity directly by recording electrical activity with the use of electrodes placed on the scalp (**Figure 1B**). Therefore, it is not dependent on blood oxygenation, and its temporal resolution is higher than other methods (Michel and Brunet, 2019). However, because electrodes are placed on the scalp, it is best suited for investigating the cerebral cortex and not deep brain structures. Classically, EEG was considered a low mobility tool because it required strict control of the movements and surroundings of participants, which limits it to the lab environment. However, in the last years, the development of new technologies has allowed for improvement in mobility by creating mobile EEG systems (Melnik et al., 2017). Such systems are a great tool to study social interactions. Even though fMRI was the first method used to perform a hyperscanning study, it is EEG that is currently the most common method used to conduct hyperscanning experiments. Its popularity comes from its most important advantage, temporal resolution. Studies of social interaction that unfold on a fast scale require a method that is sensitive to it. Until now, only EEG could account for changes in neural processing on a millisecond scale while two or more humans perform an interactive task together. The high temporal resolution allows for a more precise and different type of between-brain analysis. Another advantage of EEG for hyperscanning studies is that it is easier to measure more than just two heads at the same time, as demonstrated by Dikker et al. (2017). The relatively low price of EEG systems and the availability of mobile systems are key advantages. Early EEG hyperscanning research was conducted in the lab with full control of the environment and traditional paradigms (Babiloni et al., 2007a,b). However, with further developments, more interactive and naturalistic paradigms, like playing guitars (Lindenberger et al., 2009) or romantic kissing (Müller and Lindenberger, 2014), were proven to be feasible. In recent years, another technology, which can be combined with EEG, was developed and implemented to use in research. Namely, virtual reality (VR) (Ehinger et al., 2014; Oliveira et al., 2016; Cipresso et al., 2018) is becoming more and more present in the scientific community. It allows for creating naturalistic paradigms that are fully controlled by the experimenter. This, in combination with the EEG, might be a great tool to study social interactions.

It is worth mentioning that magnetoencephalography (MEG), a method with similar characteristic to EEG but lower mobility, was also proven to be feasible for hyperscanning measurements (Baess et al., 2012; Zhdanov et al., 2015), and it has so far been used in a study that combined it with EEG to study verbal interactions (Ahn et al., 2018). Moreover, this method was also utilized to study the interaction between mothers and children (Hirata et al., 2014; Levy et al., 2017), speaker-listener roles during natural conversation (Mandel et al., 2016), and hand kinematics in leaders and followers (Zhou et al., 2016). Recently, Boto et al. (2018) developed a mobile MEG system. Therefore, we can expect more MEG hyperscanning studies in upcoming years.

fNIRS

The last neuroimaging method that we have reviewed is functional near-infrared spectroscopy (fNIRS). Similarly to fMRI, it measures brain activity indirectly and uses the contrast between oxygenated and de-oxygenated hemoglobin, and similarly to EEG, it can best measure superficial brain areas with a low spatial resolution (1 cm) (Scholkmann et al., 2013) (**Figure 1C**). Moreover, its temporal resolution is lower than that of EEG and varies between 0.1 and 1 s (Quaresima and Ferrari, 2019). Despite these limitations, fNIRS is widely used in cognitive neuroscience for its mobility and resistance to motion artifacts. In comparison to other methods discussed here, the signal measured with fNIRS is not strongly influenced by the movement of participants. This feature allows for creating experimental paradigms that resemble real-life situations more closely than classic studies. In the case of studying social interaction that involves actions from participants, it is a critical feature that is required. The first hyperscanning fNIRS study was conducted by Funane et al. (2011) and used a simple tapping synchronization task to investigate the coherence of neural activity between two brains. Since then, many researchers adopted hyperscanning fNIRS in various types of paradigms to study social interactions (Scholkmann et al., 2013). One particularly interesting study was conducted by Nozawa et al. (2016). It involved groups of participants (four) tested in a naturalistic setting (cooperative communication). Furthermore, a recently developed fNIRS system for babies allows for investigating brain functions related to parent-child interaction (Reindl et al., 2018). Such experiments are proof of the concept that studying neural between-brain underpinnings is feasible, and it brings new insight into the understanding of human cognition.

TYPE OF ANALYSIS

The analysis and interpretation of hyperscanning data is a challenging task. First, an intra-brain type of analysis has to be adjusted to inter-brain data; alternatively, new types of analysis have to be developed. Second, it is challenging to separate inter-brain relations related to identical stimuli presented to both participants from relations that represent between-brain networks (Burgess, 2013). For the case of correlation, this involves the calculation of partial or semi-partial correlation coefficients. Similar adjustments might be done to other measures. An alternative approach compares real participant pairs with randomly selected pairs and a permutation analysis (e.g., Bilek et al., 2015). The randomly selected pairs show only the coupling due to the direct joint stimulation. Deducting

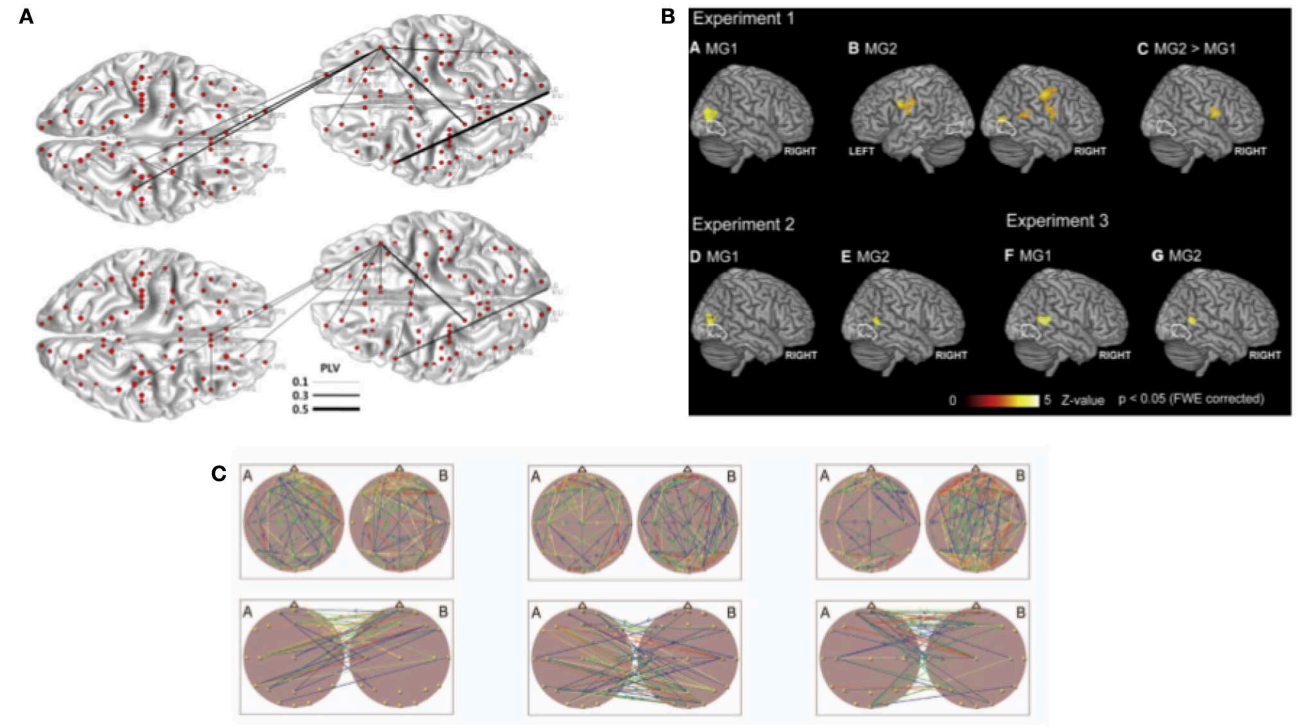


FIGURE 2 | Analysis methods used in hyperscanning to investigate between brain relationships. **(A)** From Yun et al. (2012). Phase synchrony was used as a coupling measure to investigate between-brain connections in implicit coordination task. Topography of the phase synchrony (PLV) between different regions of interest of two participants are presented for theta (4–7.5 Hz) and beta (12–30 Hz) oscillations. **(B)** From Koike et al. (2016). Between-brain synchronization estimated with correlation. **(C)** From Müller et al. (2013). Brain topography maps illustrating significant connection within and between the brains. Example of graph theory measures applied to analyze synchronization during musical improvisation on the guitar. All parts reproduced/adapted under CC licenses.

this effect from the coupling observed in the actually paired subjects uncovers the coupling of brain activity due to the genuine interaction of the partners. Overall, this is a demanding topic, and it requires precise specification of the scientific question addressed.

Furthermore, while discussing between-brain coupling measures, it is essential to mention the framework proposed by Hasson et al. (2012). It suggests that inter-brain couplings are crucial for building a shared social world. This framework builds upon research that focused on between brain couplings without the hyperscanning method. Namely, Hasson et al. (2004) presented videos to individual participants in the fMRI scanner and further analyzed between brain couplings (between all participants) related to different sections of the movie. Further, Stephens et al. (2010) used the same method to study speakers and listeners (scanning one speaker and many individual listeners to investigate the relation between the speaker and the listener). These studies were crucial for the development of the hyperscanning field and contributed to the understanding of between-brain couplings. He investigated single subjects; there are only randomly selected pairs, yet the similarities are interesting and give insight into brain functioning. Thus, the assumed control of randomly selected pairs can demonstrate interesting and insightful similarities (coupling) between participants.

The types of analyses applied to hyperscanning data can be separated into different categories. There are various coupling measures, correlation and dependence analyses, graph theory measures, and the analysis of information flow. In this section, we have discussed all these types of analyses in sequence.

Coupling/Connectivity Measures

The most common methods to estimate the strength of coupling/connectivity between brains have previously been used to study single brains. They are based on second-order measures calculated in the Fourier domain. They differ in the technical details of combining different frequencies and the kind of normalization. That is, like the phase-locking value (PLV), the phase lag index (PLI), or phase coherence have been adopted to estimate between-brain couplings. PLV measures how two signals (in case of hyperscanning coming from two different brains) are phase-locked in the observed time window. PLV is equal to 1 when phases are perfectly synchronized in a specific frequency and to 0 when they are unsynchronized. This measure was used in multiple EEG hyperscanning studies. They investigated cortical synchronization while two participants tried to imitate their hand (Dumas et al., 2010, 2011) or finger movements (Yun et al., 2012) (**Figure 2A**) during a coordinated time estimation task (Mu et al., 2016), during speaking and listening (Pérez et al., 2017), and during a cooperative decision-making task

(Hu et al., 2018). Another similar measure, also related to phase synchronization, PLI, was used in studies investigating coordinated behavior in guitar players (Lindenberger et al., 2009; Sänger et al., 2012) and also in a verbal interaction task with the use of both EEG and MEG (Ahn et al., 2018). PLV and PLI are similar measures; however, it was pointed out that PLV suffers the common source problem, and PLI does not (Aydore et al., 2013). However, for hyperscanning research, where sources are separated between brains, these measures should give the same results. Phase coherence is another method of estimating cortical synchronization within or between brains that are related to the phase of neural oscillations. It is a measure of similarity between two signals, and there is more than one way of quantifying it. Different variations of phase coherence were used in hyperscanning experiments (for detailed differences between different phase measures we recommend Thatcher, 2012). Notably, studies mentioned above investigated guitar players (Lindenberger et al., 2009; Sänger et al., 2012, 2013; Müller et al., 2013) as well as romantic partners while kissing (Müller and Lindenberger, 2014). Moreover, the latter study also estimated cross-frequency couplings between brains.

Wavelet transform coherence (WTC) is a related method to measure the coherence of two signals. It was developed to analyze the geophysical time series (Grinsted et al., 2004). However, it finds its application also in neuroscience, especially in analyzing fNIRS hyperscanning studies. Since one of the first fNIRS studies (Cui et al., 2012), it was used in the following experiments. As it is the most common method that is used to analyze inter-brain synchrony with fNIRS, it is also the most common analysis method within all hyperscanning studies. WTC was used to estimate inter-brain synchrony in paradigms studying action monitoring (Dommer et al., 2012), cooperative and competitive behaviors (Cui et al., 2012; Osaka et al., 2014; Cheng et al., 2015; Liu et al., 2016b; Wang et al., 2019), imitation (Holper et al., 2012), verbal communication (Jiang et al., 2012, 2015; Nozawa et al., 2016), non-verbal communication (Osaka et al., 2015; Hirsch et al., 2017), decision making (Tang et al., 2016; Zhang et al., 2017), coordination (Hu et al., 2017; Ikeda et al., 2017; Pan et al., 2017) , mother-child problem solving (Nguyen et al., 2019), and teaching/learning behaviors (Pan et al., 2018).

All coupling/connectivity measures mentioned in this section are measures of similarity between two neural signals coming from different brains. This similarity is interpreted as synchrony between brains in these studies. Moreover, the similarity is estimated with different methods and is always referred to as inter-brain synchrony. Such simplification of many analysis methods to address synchronization may lead to wrong interpretations of results and creates a wrong view of coherence between studies. The hyperscanning field should develop terms to distinguish between different measures of inter-brain synchrony and methods used to estimate it.

Correlation and Dependence Analysis

Another way of measuring synchrony between brains is by estimating the correlation between signals coming from two brains. Within the hyperscanning literature, we found different types of correlation measures applied to EEG, fNIRS, and fMRI data. Due to the low temporal resolution of fMRI, coupling measures discussed in the previous chapter could be used only in the very-low-frequency range, which is not typically associated with information processing in the brain. Therefore, the relation between two measured brains is often estimated with the use of linear dependence. It is not the BOLD signal itself that is used for correlation analysis, but regression model coefficients are representing activations in different tasks. These types of analyses were applied in research investigating mutual gaze, shared attention, and cooperation in the joint force production task (Saito et al., 2010; Koike et al., 2016, 2019; Abe et al., 2019) (**Figure 2B**). Correlations found in these studies were interpreted as neural synchronization between brains.

Further, two studies focused on verbal communication between participants and used correlation of BOLD activity to predict the flow of information between the sender and perceiver (Anders et al., 2011) and synchronization of brain activity between interlocutors (Spiegelhalder et al., 2014). To further extend the dependence analysis, cross-correlation in combination with ICA decomposition of the BOLD signal was used in studies focused on joint attention in participants with borderline personality disorder and healthy participants (Bilek et al., 2015, 2017). There, the cross-correlation between two brain signals was interpreted as information flow.

Correlation measures are also applied to EEG hyperscanning data. Namely, we identified studies using correlation as a measure of between brain synchronization in different paradigms. Moreover, different aspects of EEG signals were used for correlation analysis. Correlation between different frequencies (theta and alpha) was used to investigate the coordination of speech rhythm (Kawasaki et al., 2013) as well as differences between interactions between strangers and couples in alpha, beta, and gamma (Kinreich et al., 2017). Furthermore, the total independence analysis (Wen et al., 2012) was used in a study that investigated between brain synchronization in a class environment on a group of students (Dikker et al., 2017; Bevilacqua et al., 2019). This analysis was used to predict classroom dynamics and engagement.

Lastly, two fNIRS experiments applied correlation analysis to estimate synchrony between brains in tasks that required cooperation or competition between participants (Funane et al., 2011; Liu et al., 2015).

Similarly to coupling measures, the correlation and dependency analysis leave ambiguity about how to relate results from different studies due to a variety of methods applied to estimate the phenomenon of synchronization.

Graph Theory Measures

Between-brain networks can also be quantified with graph theory measures. Up to today, there are only studies that used graph theory measures on EEG hyperscanning data. Graph theory measures focus on different aspects of between brain networks. Within existing hyperscanning literature, we found studies that focused on links between brains and modularity of networks while participants performed the decision-making task (De Vico Fallani et al., 2010). Moreover, different graph theory measures were used to investigate between-brain networks in guitar players. Small-worldness of between-brain networks was enhanced during musical coordination (Sänger et al., 2012), and the topology of between-brain networks was dependent on frequency and was more regular at higher frequencies (Müller et al., 2013) (**Figure 2C**). Additionally, the directionality of between-brain networks was used to predict leaders and followers in guitar players (Sänger et al., 2013). In another study, the dimensionality of between brain networks was investigated in combat cooperation tasks (Dodel et al., 2011). All mentioned measures suggested that neural synchrony can be estimated with graph theory measures and that these measures extend our understanding of between brain networks. Few studies mentioned here give great insights into understanding neural dynamics between brains. We believe that graph theory measures are a great tool to account for the complexity of inter-brain relations. Measures like modularity, small-worldness, and directionality are bringing a new perspective into understanding neural underpinnings of dynamic social interactions. More studies should explore these measures. Additionally, more data-driven methods to define network properties are becoming more popular and can find their application in hyperscanning as well (Sporns, 2018).

Information Flow

Apart from synchrony, similarity, or network properties, one can focus hyperscanning analysis on the flow of information from one brain to another. Such studies require estimating causal links between brains. Methods that are used to determine such causal links are Granger Causality and its equivalent in a frequency domain Partial Directed Coherence (PDC). In the EEG hyperscanning literature, these methods were applied to estimate links between brains of cooperating pilots (Astolfi et al., 2011, 2012), and results suggested that causal links are stronger during increased cooperative behavior. Similarly, increased causal links between the brains of participants were found in cooperative and altruistic behaviors in decision-making tasks (De Vico Fallani et al., 2010; Ciaramidaro et al., 2018). Furthermore, one fMRI and one fNIRS study focused on causal links between brains. Schippers et al. (2010) studied such links in gesture communication with the use of fMRI and Pan et al. (2017) used fNIRS to explore causal relationships between brains of cooperating lovers. The casual links between brains can be estimated with methods that we discussed here; however, the important question of what the neural substrates that allow for information flow between brains are is yet to be answered. It is critical to understand the difference between information flow and synchronized neural activity between brains due to identical sensory input. This problem is often not addressed and left for readers to wonder how to disentangle both. Future research should focus on this aspect.

Taken together, in this section, we reviewed different methods and types of analysis that are used in the hyperscanning field. A variety of techniques and analysis suggests that hyperscanning is a new and valuable part of the cognitive neuroscience field. However, in many cases, the advantages and disadvantages of a specific method are not that obvious. Further, at least in part, we consider the growing variety of techniques used as exploratory, and it has to be investigated whether they relate to the same set of physiological processes.

COGNITIVE FUNCTIONS

Coordination and Synchronization

The investigation of interpersonal coordination of actions that includes mutual entrainment or synchronization is one of the field's most suited for hyperscanning. Simultaneously measuring the brain activity of interacting subjects allows for real-time access to the reciprocal coupling of neural processes that enable interpersonal movement synchronization within a millisecond time scale. Experimental paradigms are addressing the connection between interpersonal neural dynamics and behavioral synchronization span from minimalistic button-pressing tasks to complex naturalistic settings like joint music playing. In minimalistic tasks, different parameters, such as visual contact, feedback, and mode of synchronization (in-phase vs. antiphase), can be manipulated easily. Additionally, several studies have compared the degree of behavioral synchronization between human-human and human-computer (metronome) couples (Konvalinka et al., 2014; Mu et al., 2016; Hu et al., 2017) in order to extract the social aspect of the interaction. Such setups enable the examination of various aspects of action coordination and synchronization while controlling the effects of a shared sensory environment.

One conventional paradigm is used to study coordinated behavior and its neural underpinnings and requires participants to perform only one temporally synchronized button press after a predefined or self-time interval has passed. As a result of this, better performance was related to higher inter-brain coherence in frontal areas as well as to stronger social connectedness in the dyad (Funane et al., 2011; Mu et al., 2016; Hu et al., 2017; Pan et al., 2017).

Another paradigm used continuous tapping or finger/hand movements, allowing for additional insight into the time course and the dynamics of synchronization. Tognoli et al. (2007) found that the spontaneous transition from uncoordinated to coordinated rhythmic movements under vision went along with specific EEG rhythms in the alpha mu range at right centro-parietal sides. Dumas et al. (2010) took a between-brain approach, using the Phase Locking Value (PLV) across a variety of different frequency bands. He found that right parietal alpha mu oscillations were significantly more coupled in periods of spontaneous synchronization. Both results point toward the relevance of these patterns for the mirror neuron system. A similar paradigm also investigating alpha-band activity was used by Naeem et al. (2012a,b). However, they did not replicate Tognoli's approach but focused on broader frequency bands in the mu range in different coordination contexts (intrinsic, in-phase, and antiphase), suggesting functional discrimination of the lower (8–10 Hz) and upper (10–12 Hz) mu band (Naeem et al., 2012a). While the former seems to reflect general attentional processes, the latter is modulated by task and hemisphere: in the left hemisphere, the top mu band is present during imitation, while in the right hemisphere, it is involved in perceptual-motor discrimination. Based on this, the authors suggest a right hemispheric circuit that modulates the way the actions of others are processed concerning the desired coordination mode (Naeem et al., 2012b). In another study that focused on the directionality of interaction, the subject associated as the leader showed a characteristic suppression of frontal alpha activity, possibly representing enhanced cognitive control and planning (Konvalinka et al., 2014). Manipulating the neural synchronization between the participants with tACS, two studies directly explored the impact of phase-coupled neural oscillations on behavioral performance. Results indicated that in-phase beta but not alpha or theta stimulation across the respective motor cortices facilitated tapping entrainment (Novembre et al., 2017; Szymanski et al., 2017a). However, it was not yet shown whether this effect could be replicated in EEG studies.

In contrast to such minimalistic experimental setups, several studies applied hyperscanning in more cognitively demanding and also more naturalistic settings. Recording two subjects interacting in a finger-tapping imitation task, Holper et al. (2012) observed increased functional connectivity between two interacting brains. Social aspects modulated even unconscious fingertip movement synchronization: Yun et al. (2012) found that after having cooperated in an induced imitation task, the patterns of unconscious finger movement across two subjects became more synchronized. On a neural level, this change went along with increased theta and beta band phase locking across different regions, including the inferior frontal gyrus (IFG), anterior cingulate cortex (ACC), and ventromedial prefrontal cortex (vmPFC). In this context, the researchers associated the observed activity patterns with implicit social processing.

In a cognitively more demanding task, Lindenberger et al. (2009) investigated pairs of guitarists playing a melody together. Similarly, they reported that coordinated actions between the subjects involved oscillatory couplings between the two brains. During coordinated play onset, they found significantly increased phase synchronization between the two brains, primarily over frontal-central connections. The synchronization was exceptionally robust in low-frequency ranges, between 0.5 and 7.5 Hz, with a maximum in the theta frequency at 3.3 Hz. This result contradicted previously mentioned studies that primarily reported dominant alpha synchronization during interpersonal action coordination. The authors, however, noted that the observed couplings might merely reflect similarities in the temporal structure of the individual's perception and action. Accordingly, it is still not clear whether the neural coupling causes the effective movement coordination between the pairs. Rather than reporting specific brain areas and frequency bands, the studies mentioned above suggest that inter-brain connectivity through interpersonally coupled brain oscillations facilitates complex interpersonal action coordination.

Social aspects of action coordination, i.e., the influence of social connectedness and social character traits on synchronization performance, is another topic where multiple brain recordings provide new valuable insight. On a hormonal level, Mu et al. (2016) could show that oxytocin, intranasally administered, significantly facilitates neural synchronization in the alpha band and thus effectively supports movement coordination. Addressing the influence of the social connectedness between pairs, Pan et al. (2017) compared the performance of lovers to strangers and friends in a simple coordination task. Between the lover's brains, they indeed found a significantly increased synchronization. More specifically, they reported that the right frontoparietal network was involved in romantic processing and social cognition. At the same time, lovers also showed a significant increase in coordination performances. Since oxytocin is strongly associated with social bonding, especially in romantic relations, this result supports Mu et al. (2016), indicating the human hormone's facilitating effect on interpersonal action coordination. Applying the same simple interaction task, Hu et al. (2017) found a correlation between the prosocial inclination of the subjects and their respective inter-brain synchronization. All in all, these studies supported the evidence that social traits and the ability to synchronize in interpersonal coordination are strongly connected.

The current hyperscanning research on synchronization and coordination reports neural (synchronization) effects in various areas and frequency bands, although exploring very similar tasks. There are two groups of findings: the first group reports inter-subject neural couplings in frontal and parietal regions that are associated with better action synchronization (theta, alpha, and beta frequency). The second group of findings focuses on mechanisms that are not coupled across individuals but correspond to how a single brain processes incoming stimuli in a coordination context (Tognoli et al., 2007; Naeem et al., 2012a,b; Konvalinka et al., 2014). Interestingly, these within-brain effects were also reported at frontal and centro-parietal sides in the alpha range. They all indicate that interpersonal action synchronization is accompanied by neuronal coupling of primarily frontal and centro-parietal areas in lower frequencies. However, concerning prominent frequency bands related with movement synchronization, the reported results do not seem to be conclusive: while some effects were mainly within the alpha (10–12 Hz) and beta (∼20 Hz) range, others specifically excluded the alpha range and instead emphasized a synchronization in the theta frequency (Lindenberger et al., 2009; Yun et al., 2012; Novembre et al., 2017). Such differences in reported effects of activity in different frequencies have to be addressed in future research.

Apart from this, prosocial character traits, such as the social connectedness of the dyads, influenced the effectiveness of synchronization. However, since many of the mentioned studies had fewer than 10 pairs of subjects, more work is needed to ensure and replicate the results.

Music

Musical performances offer attractive experimental conditions since such performances combine intrapersonal action coordination and interpersonal action synchronization as well as continuous interaction. The advantages of musical settings for hyperscanning experiments are reviewed by Acquadro et al. (2016). A variety of experimental paradigms allow for the investigation of different aspects of the interaction.

To investigate that inter-brain synchronization during an interaction is not only present due to the perception of the same ecological situation, researchers assign roles to the participants to investigate if complementary roles induce asymmetric patterns of brain synchronization. Sänger et al. (2012) investigated interpersonal action coordination using EEG hyperscanning of musical leader-follower duets playing a two-voiced piece of music repeatedly. They reported within-brain phase-locking modulated by the assigned role as well as extended within- and between-brain phase coherence during phases of high musical coordination. Because of the complementary voices of the piece, the phase coherence occurs in a situation where the action and perception of the partners are not equal. Further, graph theory analyses show the presence of hyperbrain network structures. Later analysis of the same data by Sänger et al. (2013) allows for investigation of the directionality of functional connectivity between the two brains. Results show directionality as a function of the musical roles. Pan et al. (2018) recorded brain activity of learner-instructor dyads during the acquisition of two songs using one of two learning methods (part learning vs. whole learning). The study recorded fNIRS data of bilateral fronto-temporoparietal regions. Across the part-learning group, they report interpersonal brain synchronization during the learning periods, which was even able to predict the learning performance. Furthermore, Granger causality analyses show coupling directionality from instructor to learner during a particular learning phase (teaching phase). The absence of interpersonal brain synchronization correlations in the whole learning group speaks against it as a mechanism of pure shared perception since both learning groups received equal sensory input and performed comparable actions. Synchronous oscillations are a present mechanism in leader-follower musical joint action tasks, and the asynchronous nature of these signals gives rise to inter-brain synchrony partly as a mechanism of interactive task performance.

Other experimental designs investigate music without assigned roles, as this is the case in many musical contexts. While some experiments use existing music pieces, others engage in freer musical interactions like non-notated parts of songs or even improvisation. Novembre et al. (2016) used the structured properties of sheet music to manipulate familiarity and behavioral interpersonal synchronization during joint piano playing. With dyads of amateur piano players performing passages of two-voiced joint playing with congruent and incongruent instructions for a later tempo change as well as alternating knowledge about the complementary voice, they reported significant correlations between alpha suppression and congruent vs. incongruent tempo instructions in the case of the pianist being familiar with both voices of the passage. The authors concluded with the idea of alpha oscillations as neural processes regulating the balance between self-other integration and segregation, modulated by the compatibility of internal knowledge and external environmental information during joint action. After verifying EEG as a suitable method for hyperscanning in a musical context, Babiloni et al. (2011) used a hyperscanning paradigm to investigate empathy inside ensembles of musicians, playing a piece together (Babiloni et al., 2012). Alpha desynchronization in the right Brodmann area 44/45 during a video observation of their performance is positively correlated to the results of the Empathy Quotient Test score. Müller et al. (2013) investigated musical improvisation in dyads of guitarists. They analyzed intra- and inter-brain synchrony during either a phase of joint improvisation or phases where one guitarist improvised while the other listened. They reported high-frequency intra-brain connections as well as lower frequency inter-brain connections. Guitarists playing alone showed stronger out-strength than the listening guitarist in the beta range; this difference was not present during joint improvisation. Osaka et al. (2014) compared fNIRS inter-brain coherence of participants during cooperative humming of a song with or without eye contact and single humming. Results indicated enhanced wavelet transform coherence inside the right inferior frontal cortex (IFC) during the non-face-to-face condition. A further study, Osaka et al. (2015), compared the inter-brain synchrony between dyads humming or singing a song, again with or without visual contact, solo and joint. The left IFC showed increased synchronization for joint singing or humming, irrespective of the visual condition, while the right IFC showed increased synchronization specific to joint humming. The absence of synchrony in solo performances and aligned pseudo-pairs suggests the involvement of bilateral IFC in (musical) cooperation tasks. Inter-brain synchrony is a present mechanism even in more unstructured musical interactions, acting as a marker of interpersonal action coordination. Furthermore, experimental musical setups can be used for methodological analyses, as in the case of Zamm et al. (2018).

Altogether, the present results confirm musical paradigms as highly coordinative situations generating the ability to observe inter-brain synchrony as a mechanism of interpersonal action coordination with a high potential for future research.

Emotion and Affect

Emotional regulation and affect play a crucial role in various forms of social interactions, such as the willingness to undertake joint actions with peers (Lopes et al., 2005) or in different types of prosocial behavior (Twenge et al., 2007). Neuroscientific studies measuring emotion and affect based only on one participant's data lack the inter-brain connections among areas that might be involved in social behavior. In order to fill this gap, hyperscanning allows for recording inter-brain activity on emotions' onset and the simultaneous responses of interacting people.

To address the emotional component in social exchanges, several hyperscanning paradigms have been applied. Among these, setups have involved facial communication of affect (Anders et al., 2011), mother-child interactions (Hirata et al., 2014; Levy et al., 2017), and goal-seeking tasks involving cooperation and competition conditions (Pan et al., 2017). Nonetheless, due to the complexity of the setups (i.e., Hirata et al., 2014), hyperscanning studies have scarcely focused on the role of emotional regulation during joint actions (Ciaramidaro et al., 2018), leading in most cases to merely exploratory designs and vague hypotheses (Balconi and Vanutelli, 2017).

As an example of emotion processing during goal-oriented tasks, Hu et al. (2017) studied the prosocial behavior of dyads while performing a task in which participants performed coordinated and independent tasks across several trials. The authors found synchronized inter-brain activity only under the coordination task in the left middle frontal cortex (LMFC). This area has been commonly associated with memory, response inhibition, and people recognition during social interactions. Besides, Ciaramidaro et al. (2018) performed a study in which participants had the opportunity to distribute a quantity with a partner. A third participant (the observer) would judge the fairness of the distribution and decide whether to punish or not the participant who acted out the distribution. The dyads of participants involved in the exchange were composed of either human-human or computer-human. EEG data revealed higher inter-brain coherence of theta, alpha, and beta bands in the human-human condition between the observer and the receiver when the latter's action was rated as "hyper-unfair." Additionally, synchronous activity was also robust for PC-human fair interactions where the human participant received a fair reward. As the authors stated, situations with high emotional impact showed higher inter-brain synchronization.

In another experiment, Anders et al. (2011) investigated the emotional communication in romantic partners by observing the flow of information in male participants for emotional states perform by their female counterparts. They suggested that the neural activity of the perceiving partner can be successfully predicted from the neural activity of the sender's brain. This shared activation could only be found in dyads comprised of romantic partners but not in dyads of the sender and another participant different from her romantic couple. This finding suggested the development of reinforced neural paths present among sexual partners with highly emotional bindings.

Finally, some studies on emotions in social interaction have addressed the simultaneous measurement of inter-brain activity between mother and child. Levy et al. (2017), for instance, used a hyperscanning MEG setup to measure the brain-to-brain activity of mother-child dyads by exposing them to video recordings of themselves performing positive and conflictual interactions. They found gamma activity in the superior temporal sulcus (STS) in interactions with behavioral synchrony (i.e., positive interactions). STS has been amply linked to social cognition, the theory of mind, and mirroring behavior. In the same line, Hirata et al. (2014) developed a hyperscanning MEG device that enables the mother and child to see each other's facial expressions during brain activity measurement. Although mainly of an explorative kind, these studies comprised a relevant background as pioneers of experimental designs to account for emotional interaction in hyperscanning setups.

Although not extensive, these studies highlight the moderator effect that the emotional component has in inter-brain activity in two scenarios. First, the closer the relationships between participants, the higher the inter-brain synchrony as observed in romantic couples and mother-child interactions. However, there are many more possible relations between participants that have never been tested, for instance, siblings, employer-employee and seller-buyer dyads. In the future, the hyperscanning should explore other relationships between humans and emotions related to them. Second, inter-brain synchrony is higher for scenarios involving empathetic behavior, especially when these include an active emotional component. To sum up, the intensity of the emotional component modulates the synchronous neuronal activity during social interactions. Still, further research needs to be driven on this topic. For instance, the effect of well-studied emotions as stress or disgust must be investigated. This can shed light on whether the impact of negative emotional interactions induces more synchronize behavior than in the presence of emotions of a positive valence or if, instead, the modulatory effect of these might slightly depend on the sort of task.

Cooperation and Competition

Hyperscanning studies have addressed cooperative and competitive contexts under several methodological paradigms. These allow for the study of both conditions within the same setup. Therefore, participants can either cooperate or compete to achieve their goal, and meanwhile, intra- and inter-brain activity is recorded. These include, for instance, the Prisoner's dilemma task (Babiloni et al., 2007a; De Vico Fallani et al., 2010), chicken's game (Astolfi et al., 2010), time estimation (Cui et al., 2012), turn-based interaction disk games (Liu et al., 2015, 2016b), Jenga (Liu et al., 2016a), or pong-game (Sinha et al., 2016).

Concerning intra-brain activity, hyperscanning studies reveal some commonalities of activation around the prefrontal cortex (PFC). For instance, during a Prisoner's Dilemma task, Babiloni et al. (2007a) found that mPFC is active during all the conditions (i.e., cooperation, defect, and tit-for-tat). In contrast, ACC is only activated when participants defect. In general, the global integration of brain areas was higher under the competitive condition than in cooperation and tit-for-tat. This is in line with findings by Astolfi et al. (2010) in which defect and tit-for-tat conditions obtained higher activity than for the cooperative condition in beta-band EEG recordings. mPFC has been generally related to social interaction supporting the constant activation observed during all conditions. On the other side, ACC has been linked to the theory of mind, indicating that an extra effort is needed to predict the opponent's behavior under competitive interactions. In another scenario, Liu et al. (2015, 2016b) performed a turn-based interaction in a computerized two-person game. Participants took turns to be either a builder or a helper/obstructer partner; brain activity was recorded using fNIRS. They found significant activation in rIFG in builders during the cooperation condition but not when their partners were competing. A similar set-up was used by Liu et al. (2016a) in which a significantly higher activity was found in the obstructors' rIFG area. However, in both studies, no effect was found for helpers; i.e., no "cooperated effect" was revealed. rIFG has been linked to empathy and intention understanding during interpersonal interactions. In this sense, results showed a need for higher empathy when guidance is necessary to achieve a common goal. On the other side, when it comes to hinder other's performance, the understanding of an opponent's intentions plays a crucial role.

On an inter-brain level, the activation in PFC seems to be modulated by the condition and nature of the task. For instance, in the aforementioned setup, Liu et al. (2015, 2016b) found a significant inter-brain synchrony only in builder obstructor pairs. Additionally, Liu et al. (2016a) observed active inter-brain synchrony in the posterior region of the right middle and superior frontal gyrus, particularly Brodmann area 8 (BA8), during cooperative and obstructive interaction (but not in the parallel game and talking condition). Inter-brain synchrony was also observed only during cooperative interaction in the dorsomedial prefrontal cortex (dmPFC), particularly in Brodmann area 9 (BA9). Since participants are performing a joint activity, motor execution has to be synchronized. This is in line with previous findings linking PFC with functions as planning and motor execution. On the contrary, other studies reported stronger synchronized inter-brain activity in cooperative contexts. As shown when De Vico Fallani et al. (2010) and Babiloni et al. (2007a) performed Prisoner's dilemma setups, hyper brain networks in competitive brains have fewer links and have overall higher modularity than in tit-for-tat and cooperative couples. Furthermore, Cui et al. (2012) found increased coherence between signals measured over the right superior frontal cortices between two brains in cooperative and not during competitive behavior. Supporting these findings, Sinha et al. (2016) reported significantly higher inter-brain synchrony between the subjects when they cooperated as compared to the competitive scenario. Additionally, they found that inter-brain synchrony was enhanced considerably when the subjects were physically separated, i.e., they cooperated via an intranet network. This is in contrast with Liu et al.'s (2015, 2016a,b) findings of synchronized activity in dmPFC in competitive contexts. This might be because different set-ups require synchronized activation under different conditions. For instance, a task like the prisoner dilemma needs a higher understanding of other's intentions when participants decide to cooperate.

All in all, hyperscanning studies confirm previous findings on the crucial role of dmPFC in collective behavior. However, the strength of this synchronized activation in dmPFC depends not only on the condition (i.e., cooperation and competition) but also on the specific kind of task as well. For instance, tasks like turn-taking games (e.g., Jenga) that require the prediction of the opponent's actions demand a higher level of the theory of mind processing. On the other hand, tasks like the prisoner's dilemma imply empathy/theory of mind during the cooperative scenarios, and these differences are also reflected in between-brain analysis. With further development of mobile neuroimaging methods, studying cooperative and competitive situations might be possible in more real-life situations. For example, we can imagine using sports games like football or basketball, where players cooperate and compete at the same time with other players. It would be interesting to see whether results from experimental hyperscanning scale to real-life cooperative and competitive situations.

Games and Decision Making

Overall studies in the field of games and decision making have shown that their neural underpinnings involve a network of regions. They range from the medial frontal cortex (MFC), superior temporal sulcus (STS), and to the temporoparietal junction (TPJ). Throughout the last years of research in the field of interactive decision making in games, a specialization of focus took place, as the first studies focused on areas active in simple games, such as game theory. However, the first investigations to test the neural basis of social interaction used the game theory, as it allows us to define a social situation in which one may lose or profit. Babiloni et al. (2007a) demonstrated that cooperative social interaction activates the reward circuitry. Non-cooperative behavior, in contrast, does not. Their findings suggest a strong activation of the ACC and the cingulate motor area (CMA). The results point out the importance of the ACC, especially for leaders. In their case, it was the person who plays the first card on the deck.

Besides, Babiloni et al. (2006) presented EEG hyperscanning as a new and valid methodology to address the brain activity of a group during real-life social interaction, the "spirit of the group." Building upon the findings from Babiloni et al. (2006), they addressed social interaction during a game. The aim this time was to measure the neural activity of different brains simultaneously, particularly neural processes generated by social cooperation or competition. The results are similar, and they also provided evidence for the ACC and the CMA to be maximally active (Babiloni et al., 2007b). One other early experiment in the field of decision making was performed by Tomlin (2006). They investigated the impact of personal and impersonal situations by using fMRI hyperscanning. Their findings were in line with the results by Babiloni et al. (2007a), as the dorsal anterior cingulate cortex responded strongly to their set up. Furthermore, cingulate and paracingulate cortices appear to contribute to social cognition and decision-making.

Further, Tomlin (2006) added the possibility that other variables in the social domain may impact outcomes in this area, like the belief in "me" or "not me." Also, Yun et al. (2008) studied social decision making by using the Ultimatum Game, as the experimental model offers the estimation of e.g., fairness or mind-reading, which has been used before as well (Sanfey, 2003). They also mentioned, as other authors have done, the umbrella term "theory of mind," showing how wide the topic can be interpreted. Their results suggested high-frequency oscillations in frontocentral regions, indicating that social interaction is closely related to this area. Investigating the effect of gender in cooperative and non-cooperative situations, Cheng et al. (2015) used fNIRS and revealed that task-related coherence in brain activity. This was evident in regions of the frontal cortex, especially when opposite-sex partners are cooperating. The last study to mention here is the one by Zhang et al. (2017), as they provided an overview of research from the last years and focused on another variable deception. In their study, they used fNIRS hyperscanning to measure pairs of participants in a two-person gambling card game simultaneously. Their findings provided higher TPJ activation in deceptive acts compared to honest ones. Further, they assume that STS may play a critical role in spontaneous deception. Decision making in games offers a well-controlled environment to investigate decision making. Future research has to uncover the precise influence of a known and not know partner, and the differentiation between cooperation and competition. Furthermore, influences like facial expression or gestures are worth considering.

Action Representation and Joint Attention

Whenever we socially interact with others, we have to coordinate our actions with those of our partners precisely. For successful joint action, we need to understand our partner's intentions and combine it with our action plan, always anticipating, attending, and adapting. In this context, joint attention provides the basis for shared awareness of common objects and goals, which is required to join our actions with others effectively. When studying neural mechanisms underlying these cognitive abilities, hyperscanning research provides new opportunities to investigate the intra- and inter-brain effects that accompany joint action. Setups range from pure natural eye-to-eye contact and mutual visual search to more demanding joint musical performance.

Considering mutual gaze as the communicative context in which joint attention is initiated, Hirsch et al. (2017) investigated the neural effects of natural eye-to-eye contact via fNIRS. Comparing "online" interactive eye-to-eye-contact with an "offline" non-interactive eye-to-picture condition, they reported a broad neural network reacting sensitive to interactive mutual gaze: during online eye-to-eye contact, the hemodynamic signals of the left frontal (pre- and supplementary Motor Cortex) the and temporal-parietal regions displayed a higher functional connectivity within brains as well as increased synchronization between brains. This network vastly overlaps with regions associated with language perception and interpretation (i.e., Broca's and Wernicke's regions). Due the this, Hirsch and his team supposed that natural eye-to-eye-contact actively incorporates face-to-language processing. The cross-brain coherence observed in these areas supports this claim, indicating that the rapid online exchange of information between the brains that enables language processing is also communicatively active during mutual gaze.

Further research investigating mutual gaze has used similar experimental paradigms: they observed the brain activity of two subjects interacting in a non-verbal joint attention task (Saito et al., 2010; Lachat et al., 2012; Koike et al., 2016). Here, subjects had to mutually attend target objects either by following the partner's gaze, by self-initiating the common gaze direction, or by following an external cue. In the hyperscanning fMRI study of Saito et al. (2010), during moments of shared attention, paired subjects showed significantly higher inter-brain correlations in the IFG. They therefore concluded that observed inter-brain synchronization in the right IFG facilitates the formation of shared representations, enabling the incorporation of shared intentions by internalizing the other's intentions.

These findings closely relate to the reports of an extended fMRI study by Koike et al. (2016). In this experiment, the research additionally examined the eye-blink synchronization between the subjects, considering them as an index of joint attention. Alternating between mutual gaze and joint attention tasks, dyads displayed increased synchronization of eye-blinks and right IFG activity when they had been previously engaged in a joint attention task. The researchers take this as an indication that the inter-personal neural synchronization through joint attention can be learned, and therefore, be maintained in the social memory. Similar to Saito et al. (2010), the study also reported significant inter-brain synchronization in the right IFG in the context of initiating as well as responding to joint attention. This synchronized activity also correlated positively with enhanced eye-blink synchronization. Importantly, in a video control condition, where participants did not see their partners as a live recording, the right IFG showed no activity. From these results, the study inferred that the right IFG acted as an interface between the self and the other; it is thus thought to coordinate constant shifts between central-executive and default-mode networks, moving attention between oneself and the partner. This fits well with Saito et al. (2010); they associated the synchronized activity of the right IFG with the formation of shared representations between subjects.

Applying dual EEG to compare the neural activity of the socially driven vs. color-driven gaze direction, Lachat et al. (2012) based their research on different brain oscillations. They focused on frequency bands around 10 Hz over parieto-occipital and centro-parietal since this activity is generally associated with social coordination abilities. As previously expected, they found an attenuation of left-hemispheric alpha and mu rhythms by joint attention. This modulatory effect, however, was characteristic for the mutually directed gaze in general, independent of the type of instruction, i.e., whether it was socially or color driven. The researchers interpreted this suppression of the alpha mu rhythm as an indication for an "attention mirroring system," which allows subjects to orient their attention jointly. The left lateralization of this alpha mu attenuation contradicts previous research, where neural effects of social interaction are predominantly reported in the right hemisphere (Saito et al., 2010; Dumas et al., 2012; Koike et al., 2016; Novembre et al., 2016).

In contrast to these mutual gaze experiments, Szymanski et al. (2017b) compared individual performance with a joint performance during a visual search task. Here, the interaction between subjects was much more natural since verbal, gestural, and tactile communication could be used freely. The researchers tried to relate within- and between-brain neural dynamics to their respective team performance. Indeed, their results indicated that the overall team performance increased with intra- and inter-brain phase synchronization, especially in lower frequencies at frontal sites. Thus, local as well as between-brain phase synchronization were considered as supportive factors for joint attention performance.

Beyond joint attention, the question of how two persons coordinate their actions with one another is subject of hyperscanning paradigms. Following the notion of co-representation (Sebanz et al., 2003), humans form an internal representation of another person's actions through common coding and mirror neuron mechanisms. This representation helps to adjust their actions in favor of a (joint) goal. However, the nature of human interactions is divers; relationships can be symmetric or complementary and emerge spontaneously or be predefined by the type of social situation. The question of how the representation of the self's and other's actions are modulated in these different contexts was the main subject of the studies discussed in the following section.

Ménoret et al. (2014) investigated changes in electrophysiological patterns when we do not only observe an action but also co-act with our partner by performing a complementary task. They found that co-acting led to stronger movement-related beta suppression and more negative movement-related potentials at frontal sides in observers. This implies that co-acting goes along with a more intense representation of the other's action compared to mere observation. Sebanz et al. (2006) led two people to perform a go/nogo task alone or as a pair sitting side by side. Each subject reacted to a different color cue, while a task-irrelevant stimulus pointed to a side either compatible or not compatible with the side on which the participant who was in turn to press the button was seated. Longer reaction times in the incompatible condition and a stronger Nogo P3 component at frontal and central electrodes in the group condition can be interpreted as a consequence of co-representing the partner's actions and the need to suppress own action-tendencies. Both Ménoret's and Sebanz's findings are in line with the concept of co-representation, indicating that observed as well as expected actions activate the according movement-related mechanisms within partners.

The relation between the anticipation of a partner's actions and dynamical entrainment was subject of Novembre et al. (2016). In his paradigm, subjects either familiar or unfamiliar with the partner's notes played a short melody together while tempo instructions were manipulated. Results showed that subjects unfamiliar with their partner's part acted more adaptively. On a neural level, modulations of alpha power at right centro-posterior sides were found: when subjects knew their partner's part, an incongruent tempo between the pianists led to a power increase, while good entrainment (based on congruent tempo instructions) led to an alpha power decrease. This allows for the interpretation that alpha power modulates processes of self-other integration and segregation. While the former is present when the tempo instructions match, the latter is observed when the tempo of the partner must be ignored in order to follow the instructions.

Dumas et al. (2012) aimed at distinguishing correlates of self-other-agency in a hand gesture imitation paradigm. Contrasting analyses across a broad frequency range (0–48 Hz) were used to extract differences between the conditions "not moving and not observing," "observing gestures passively," "performing gestures alone," "induced imitation," and "spontaneous imitation." In induced imitation, the roles of model and follower were predefined by the experimenter, whereas they were established by the subjects in the spontaneous imitation condition. In the conditions where subjects performed and observed and performed gestures, a decrease in alpha mu power was observed over sensorimotor areas, including the temporal-parietal junction (TPJ). Hence alpha mu desynchronization might be a marker of action-perception couplings. When subjects were primarily observing the action, passively or as imitators, theta power increased. In the spontaneous condition, gamma was boosted across parietal regions, possibly representing the shared agency. The activation in parietal areas can be seen as a hint endorsing the relevance of TPJ for the agency and social interaction. Dumas et al. (2010) found an increased between brain phase locking in the alpha mu range during spontaneous synchronization.

While Dumas investigated random gestures, Schippers et al. (2010) addressed meaningful gestures used in a charade game. Gesturers, guessers, and control subjects that observed the gestures without guessing took turns in an fMRI scanner. Intending to find correlates of the mirror neuron system and mentalizing system, the researchers calculated the Granger causality between brains. The results supported the relevance of the mirror neuron system for action representation, as the activity in the parietal region (associated with the mirror neuron system) of the gesturer predicted activity in the mirror neuron system and vmPFC (mentalizing system) of the observer. However, the involvement of the vmPFC was both statistically and theoretically less well-funded than the mirror neuron system.

Based on these hyperscanning findings on joint attention and action representation, the relevance of the mirror neuron system and between-brain connectivity in joint action representation gained further interest. This was shown directly by inter-brain coherence (Dumas et al., 2010; Schippers et al., 2010) as well as indirectly utilizing observation-related potentials and oscillatory patterns elicited during joint action contexts (Sebanz et al., 2006; Ménoret et al., 2014). There were also power modulations related to different modes of (joint) action found across a vast range of frequencies and regions, with alpha mu being the most prominent one, perhaps representing action-perception couplings (Dumas et al., 2012; Lachat et al., 2012; Novembre et al., 2016). When it comes to fluently segregating and integrating self- and other-related information during interpersonal coordination of actions, inter-brain synchronization seems to play a pivotal role.

To further validate the proposed hypotheses ascribed to these effects, repeating experiments in combination with different neuroimaging techniques might be useful to overcome the limitations each method has. This would also increase comparability across setups and thus allow for a complete picture and a better interpretation of the findings.

Over Two Heads

Naturalistic settings are attractive conditions for studying human interaction because, in such settings, interaction occurs without the intervention of the researcher, increasing the ecological validity of the findings. In the last years, researchers have begun to extend hyperscanning research toward multi-subject setups to increase the natural component of social interactions. Early group studies were EEG hyperscanning of four participants playing the Italian card game "Tressette" (Babiloni et al., 2006, 2007b; Astolfi et al., 2010).

There are a variety of reasons for conducting experiments with a multi-subject design with different ideas of making the studied interaction more natural. Social behavior only evolves in the presence of other people, often groups. The presence of other people might enhance individual task performances (Wahn et al., 2018, 2019). In a dyadic setup, interactions might quickly become predictable. Extending the dyadic setup to larger groups may increase the complexity of the interaction due to the actions influencing more individuals generating more possible outcomes. Competition becomes more competitive, and cooperation tasks might become more complex, requiring better interaction from all members of the group. In the context of musical group performances, the structure of a leader and a follower often does no longer exist, playing in an ensemble requires continuous interaction of all members (Babiloni et al., 2011, 2012). In general, the roles of the participants in the interaction become less discrete. This is similar to many social interactions in daily life. Researchers use these properties for two kinds of experimental designs. Some apply findings, conducted from earlier experiments using dyadic design, to a group design to investigate whether these findings still hold under the more natural conditions. Other publications claim that the effects they want to observe can only be present inside a group interaction. Hyperscanning thus allows for investigation of effects that are only present inside large groups like, for example, classrooms, allowing different and new research questions. Such experiments therefore observe social behavior inside a social setting. There thus exist two main categories of current multi-subject hyperscanning research.

Multi-subject hyperscanning experiments can be used to confirm results derived from less complex social situations in a more natural setting. Gevins et al. (2012) generate a measure to distinguish subjects under the influence of alcohol from others by their EEG data. The measure was derived from EEG data, recorded from non-interactive task performance. This measure is then applied to EEG data, simultaneously recorded from each participant of a cocktail party, and still correctly discriminates subjects under the influence of alcohol or placebos. Multi-subject Hyperscanning experiments hence offer a potential method for investigations regarding social behavior.

For other researchers, hyperscanning offers a new opportunity to precisely record human group interactions to investigate social dynamics. Dikker et al. (2017) investigated brain synchrony from a class of 12 high school students over one semester during regular classroom activities. The results suggested that the individuals that are less engaged with the classroom setting show lower brain-to-brain synchrony than the rest of the group. Nozawa et al. (2016) investigated brain synchrony inside 12 groups of four members playing a word chain game under a cooperative condition, reporting frontopolar interpersonal neural synchronization by natural and unstructured verbal communication. Results like these suggest that multi-subject hyperscanning experiments can also be conducted to observe the effects of social interactions directly.

Speech and Communication

Speech is one of the most crucial aspects of social interactions in humans. The majority of human-human interactions involve verbal communication. Consequently, it is vital to study it with the hyperscanning method to understand the neural underpinnings of verbal communication. The first study that focused on verbal communication compared inter-brain synchrony between face-to-face and back-to-back dialog and monolog situations (Jiang et al., 2012). They found increased inter-brain synchrony between partners in face-to-face dialog but not in the other type of communications. This result suggests that interactive paradigms are required to observe inter-brain synchrony and that hyperscanning is a valid method to measure it. Similarly, greater inter-brain coherence between partners was found in interactive than non-interactive object-naming and description task (Hirsch et al., 2018) as well as for match over mismatch sentences (İşbilir et al., 2016). In another study, Kawasaki et al. (2013) compared the coordination of speech rhythm between human-human and human-machine dyads. Their results, higher between brain synchronization in theta and alpha bands in temporal and lateral-parietal regions, further corroborate that interaction between communicating humans is related to higher inter-brain synchrony. Moreover, when bigger groups (four participants at once) were studied during cooperative communication, frontopolar inter-brain synchronization was found (Nozawa et al., 2016). Inter-brain synchrony and coherence effects could be merely an epiphenomenon of auditory processing. This question was addressed by Pérez et al. (2017). He pointed out that speech-to-brain synchronization is mediated by low-level auditory mechanisms. Of note is the fact that it is the interactive process, however, that plays a crucial role in the inter-brain synchronization. This evidence gives strong support to claim that interaction between participants of a dialog is related to inter-brain synchrony.

Conveying information between interlocutors is a fundamental facet of human communication, especially between teachers and students. Such a scenario was studied by Holper et al. (2013). A correlation analysis between students and teachers showed that in successful educational dialogs, the brain activity of students and teachers synchronizes. As it is first and the only one study focused on the teacher/student inter-brain synchrony, more research is required to understand this phenomenon.

In general, we believe that studying speech and communication requires interaction between participants, and therefore hyperscanning is the best method to understand the neural basis of speech and communication. However, artifacts generated by speech are difficult to remove, and this limitation has to be addressed appropriately.

Intervention Methods

Intervention methods are especially appealing because properties of the object of investigation are directly manipulated: the activity of specific neural populations in the brain is up- or downregulated by physiological or pharmacological means. This facilitates changes in behavior to distinct neural processes of social interaction. Mu et al. (2016) applied EEG hyperscanning and studied the effects of oxytocin in males on the performance in a reciprocal synchronization task. The task was to synchronize a button press (varying delay in the second range) with the interaction partner or a computer. In contrast, Novembre et al. (2017) applied transcranial alternating current stimulation targeting the motor cortices of participants of each dyad. The authors compared behavioral measures for differences between in-phase and out-of-phase stimulation across subjects in a joint tapping paradigm. Similarly, Szymanski et al. (2017a) targeted the effects of same-phase-same-frequency hyper-tACS on the performance of participants in a joint drumming experiment. All three studies used synchronicity of behavior as a behavioral measure. Significant effects of Oxytocin on the mean alpha-band inter-brain PLV of posterior and central electrodes of males were found only for the social condition. However, most electrodes showed significant differences in this condition. In contrast, if participants synchronized their behavior to a computer, the difference between the treatment group and control was absent (Mu et al., 2016). Results from named tACS-studies show deviating results. Novembre et al. (2017) found higher interpersonal tapping synchrony for in-phase stimulation only for stimulation at 20 Hz. In contrast, Szymanski et al. (2017a) did not find meaningful effects of in-phase stimulation on behavior. Future research may profit from the increase of understanding of intervention methods and theoretical grounding of expected and observed effects. It is challenging to draw a conclusion with only three studies. Therefore, the understanding of inter-brain relations might be fostered by an increased amount of studies applying different intervention methods in combination with hyperscanning.

CONCLUSIONS

Taken together, in this review, we first presented methods that are used to measure the brain activity of two or more participants simultaneously. We discussed their advantages and disadvantages for studying different aspects of social interaction. Further, we reviewed the analysis methods that are used to study between brain networks. We listed different types of analyses that can contribute to various aspects of our understanding of the social brain. In the final section, we presented results of hyperscanning studies performed in the last two decades that focused on diverse cognitive functions and their neural underpinnings.

All these methods, analysis, and experimental results are in line with the call for a more ecologically valid way of studying the social brain (Hari and Kujala, 2009; Hasson et al., 2012; Hari et al., 2013; Schilbach et al., 2013; Redcay and Schilbach, 2019; Konvalinka and Roepstorff, 2012). This call is present since the last decade and suggests that we need more interactive paradigms and neuroimaging data coming from more than one brain to understand the human brain and its social nature fully. Social interactions are a fundamental part of every human being's life, and studying them is indispensable for neuroscience. The previously challenging idea of hyperscanning research was addressed in last years in multiple ways. With our review, we presented an overview and the results of this effort. Taken together, the results of different hyperscanning studies presented support the claim that hyperscanning is a useful and promising method to study social interaction. Inter-brain synchrony appears to be related to the interaction between participants. Without simultaneous measurements of more than one brain, it would not be possible to explore neural underpinnings of social interaction. However, as the field of hyperscanning is young, and in most cases, only exploratory, more research is required to understand all principles and neural basis of human social behavior. Furthermore, presented here results may give rise to a more extended view on studying the human brain. Namely, the fact that brains of participants synchronized with each other may raise a question of whether studying higher cognitive functions should include more participants to understand the human brain fully.

In sum, with the evidence presented in this review, we tried to give an informed overview of the field and point out future avenues of research to foster insights into the interacting mind/brain.

AUTHOR CONTRIBUTIONS

AC and PK organized and supervised the reading club at the University of Osnabrück to summarize and discuss hyperscanning literature. AC, SE, AL, DM, MG, SS, FS, ZR, and PK summarized all papers discussed in the review and wrote the review. AC and PK drafted and revised of the manuscript.

ACKNOWLEDGMENTS

We gratefully acknowledge the support by the European Commission Horizon H2020-FETPROACT-2014 641321-socSMCs, DFG-funded Research Training Group Situated Cognition (GRK 2185/1) and the Deutsche Forschungsgemeinschaft (DFG) Open Access Publishing Fund of Osnabrück University. Moreover, we would like to thank Charlotte Demandt, Anastasia Mukhina, Maria Pershina, and Hannah Westerhagen for participation and contribution to the reading club that was organized at the University of Osnabrück to summarize and discuss hyperscanning literature.

REFERENCES

İşbilir, E., Cummins, F., and Ayaz, H. (2016). "Investigating brain-brain interactions of a dyad using fNIR hyperscanning during joint sentence reading task," in *3rd International Symposium on Brain Cognitive Science* (Istanbul).

Abe, M. O., Koike, T., Okazaki, S., Sugawara, S. K., Takahashi, K., Watanabe, K., et al. (2019). Neural correlates of online cooperation during joint force production. *Neuroimage* 191, 150–161. doi: 10.1016/j.neuroimage.2019.02.003

Acquadro, M. A. S., Congedo, M., and De Riddeer, D. (2016). Music Performance as an experimental approach to hyperscanning studies. *Front. Hum. Neurosci.* 10:242. doi: 10.3389/fnhum.2016.00242

Ahn, S., Cho, H., Kwon, M., Kim, K., Kwon, H., Kim, B. S., et al. (2018). Interbrain phase synchronization during turn-taking verbal interaction-a hyperscanning study using simultaneous EEG/MEG: synchronization during turn-taking verbal interaction. *Hum. Brain Mapp.* 39, 171–188. doi: 10.1002/hbm.23834

Anders, S., Heinzle, J., Weiskopf, N., Ethofer, T., and Haynes, J.-D. (2011). Flow of affective information between communicating brains. *Neuroimage* 54, 439–446. doi: 10.1016/j.neuroimage.2010.07.004

Astolfi, L., Cincotti, F., Mattia, D., De Vico Fallani, F., Salinari, S., Vecchiato, G., et al. (2010). "Imaging the social brain: multi-subjects EEG recordings during the chicken's game," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology* (Buenos Aires: IEEE), 1734–1737.

Astolfi, L., Toppi, J., Borghini, G., Vecchiato, G., He, E. J., Roy, A., et al. (2012). "Cortical activity and functional hyperconnectivity by simultaneous EEG recordings from interacting couples of professional pilots," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (San Diego, CA: IEEE), 4752–4755.

Astolfi, L., Toppi, J., Borghini, G., Vecchiato, G., Isabella, R., De Vico Fallani, F., et al. (2011). "Study of the functional hyperconnectivity between couples of pilots during flight simulation: An EEG hyperscanning study," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Boston, MA: IEEE), 2338–2341.

Aydore, S., Pantazis, D., and Leahy, R. M. (2013). A note on the phase locking value and its properties. *Neuroimage* 74, 231–244. doi: 10.1016/j.neuroimage.2013.02.008

Babiloni, C., Buffo, P., Vecchio, F., Marzano, N., Del Percio, C., Spada, D., et al. (2012). Brains "in concert": frontal oscillatory alpha rhythms and empathy in professional musicians. *Neuroimage* 60, 105–116. doi: 10.1016/j.neuroimage.2011.12.008

Babiloni, C., Vecchio, F., Infarinato, F., Buffo, P., Marzano, N., Spada, D., et al. (2011). Simultaneous recording of electroencephalographic data in musicians playing in ensemble. *Cortex* 47, 1082–1090. doi: 10.1016/j.cortex.2011.05.006

Babiloni, F., Astolfi, L., Cincotti, F., Mattia, D., Tocci, A., Tarantino, A., et al. (2007a). "Cortical activity and connectivity of human brain during the Prisoner's dilemma: an EEG hyperscanning study," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Lyon: IEEE), 4953–4956.

Babiloni, F., Cincotti, F., Mattia, D., De Vico Fallani, F., Tocci, A., Bianchi, L., et al. (2007b). "High resolution EEG hyperscanning During a card game," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Lyon: IEEE), 4957–4960.

Babiloni, F., Cincotti, F., Mattia, D., Mattiocco, M., Fallani, F. D. V., Tocci, A., et al. (2006). "Hypermethods for EEG hyperscanning," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society* (New York, NY: IEEE), 3666–3669.

Baess, P., Zhdanov, A., Mandel, A., Parkkonen, L., Hirvenkari, L., Mäkelä, J. P., et al. (2012). MEG dual scanning: a procedure to study real-time auditory interaction between two persons. *Front. Hum. Neurosci.* 6:83. doi: 10.3389/fnhum.2012.00083

Baillet, S. (2017). Magnetoencephalography for brain electrophysiology and imaging. *Nat. Neurosci.* 20, 327–339. doi: 10.1038/nn.4504

Balconi, M., and Vanutelli, M. E. (2017). Cooperation and competition with hyperscanning methods: review and future application to emotion domain. *Front. Comput. Neurosci.* 11:86. doi: 10.3389/fncom.2017.00086

Bevilacqua, D., Davidesco, I., Wan, L., Chaloner, K., Rowland, J., Ding, M., et al. (2019). Brain-to-brain synchrony and learning outcomes vary by student-teacher dynamics: evidence from a real-world classroom electroencephalography study. *J. Cogn. Neurosci.* 31, 401–411. doi: 10.1162/jocn_a_01274

Bilek, E., Ruf, M., Schäfer, A., Akdeniz, C., Calhoun, V. D., Schmahl, C., et al. (2015). Information flow between interacting human brains: identification, validation, and relationship to social expertise. *Proc. Natl. Acad. Sci. U.S.A.* 112, 5207–5212. doi: 10.1073/pnas.1421831112

Bilek, E., Stößel, G., Schäfer, A., Clement, L., Ruf, M., Robnik, L., et al. (2017). State-ependent cross-brain information flow in borderline personality disorder. *JAMA Psychiatry* 74:949. doi: 10.1001/jamapsychiatry.2017.1682

Boto, E., Holmes, N., Leggett, J., Roberts, G., Shah, V., Meyer, S. S., et al. (2018). Moving magnetoencephalography towards real-world applications with a wearable system. *Nature* 555, 657–661. doi: 10.1038/nature26147

Burgess, A. P. (2013). On the interpretation of synchronization in eeg hyperscanning studies: a cautionary note. *Front. Hum. Neurosci.* 7:881. doi: 10.3389/fnhum.2013.00881

Cheng, X., Li, X., and Hu, Y. (2015). Synchronous brain activity during cooperative exchange depends on gender of partner: a fNIRS-based hyperscanning study: synchronous brain activities. *Hum. Brain Mapp.* 36, 2039–2048. doi: 10.1002/hbm.22754

Ciaramidaro, A., Toppi, J., Casper, C., Freitag, C. M., Siniatchkin, M., and Astolfi, L. (2018). Multiple-brain connectivity during third party punishment: an EEG hyperscanning study. *Sci. Rep.* 8:6822. doi: 10.1038/s41598-018-24416-w

Cipresso, P., Giglioli, I. A. C., Raya, M. A., and Riva, G. (2018). The past, present, and future of virtual and augmented reality research: a network and cluster analysis of the literature. *Front. Psychol.* 9:2086. doi: 10.3389/fpsyg.2018.02086

Cui, X., Bryant, D. M., and Reiss, A. L. (2012). NIRS-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation. *Neuroimage* 59, 2430–2437. doi: 10.1016/j.neuroimage.2011.09.003

De Vico Fallani, F., Nicosia, V., Sinatra, R., Astolfi, L., Cincotti, F., Mattia, D., et al. (2010). Defecting or Not defecting: how to "read" human behavior during cooperative games by EEG measurements. *PLoS ONE* 5:e14187. doi: 10.1371/journal.pone.0014187

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., et al. (2017). Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Curr. Biol.* 27, 1375–1380. doi: 10.1016/j.cub.2017.04.002

Dodel, S., Cohn, J., Mersmann, J., Luu, P., Forsythe, C., and Jirsa, V. (2011). "Brain signatures of team performance," in *Foundations of Augmented Cognition. Directing the Future of Adaptive Systems*, Vol. 6780, eds D. D. Schmorrow and C. M. Fidopiastis (Berlin; Heidelberg: Springer Berlin Heidelberg), 288–297.

Dommer, L., Jäger, N., Scholkmann, F., Wolf, M., and Holper, L. (2012). Between-brain coherence during joint n-back task performance: a two-person functional near-infrared spectroscopy study. *Behav. Brain Res.* 234, 212–222. doi: 10.1016/j.bbr.2012.06.024

Dumas, G., Lachat, F., Martinerie, J., Nadel, J., and George, N. (2011). From social behaviour to brain synchronization: review and perspectives in hyperscanning. *IRBM* 32, 48–53. doi: 10.1016/j.irbm.2011.01.002

Dumas, G., Martinerie, J., Soussignan, R., and Nadel, J. (2012). Does the brain know who is at the origin of what in an imitative interaction? *Front. Hum. Neurosci.* 6:128. doi: 10.3389/fnhum.2012.00128

Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., and Garnero, L. (2010). Inter-brain synchronization during social interaction. *PLoS ONE* 5:e12166. doi: 10.1371/journal.pone.0012166

Ehinger, B. V., Fischer, P., Gert, A. L., Kaufhold, L., Weber, F., Pipa, G., et al. (2014). Kinesthetic and vestibular information modulate alpha activity during spatial navigation: a mobile EEG study. *Front. Hum. Neurosci.* 8:71. doi: 10.3389/fnhum.2014.00071

Eisenberger, N. I. (2003). Does rejection hurt? an fMRI study of social exclusion. *Science* 302, 290–292. doi: 10.1126/science.1089134

Ferrari, M., and Quaresima, V. (2012). A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application. *Neuroimage* 63, 921–935. doi: 10.1016/j.neuroimage.2012.03.049

Funane, T., Kiguchi, M., Atsumori, H., Sato, H., Kubota, K., and Koizumi, H. (2011). Synchronous activity of two people's prefrontal cortices during a cooperative task measured by simultaneous near-infrared spectroscopy. *J. Biomed. Opt.* 16:077011. doi: 10.1117/1.3602853

Gevins, A., Chan, C. S., and Sam-Vargas, L. (2012). Towards measuring brain function on groups of people in the real world. *PLoS ONE* 7:e44676. doi: 10.1371/journal.pone.0044676

Glover, G. H. (2011). Overview of functional magnetic resonance imaging. *Neurosurg. Clin. N. Am.* 22, 133–139. doi: 10.1016/j.nec.2010.11.001

Grinsted, A., Moore, J. C., and Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlin. Process. Geophys.* 11, 561–566. doi: 10.5194/npg-11-561-2004

Hari, R., Himberg, T., Nummenmaa, L., Hämäläinen, M., and Parkkonen, L. (2013). Synchrony of brains and bodies during implicit interpersonal interaction. *Trends Cogn. Sci.* 17, 105–106. doi: 10.1016/j.tics.2013.01.003

Hari, R., and Kujala, M. V. (2009). Brain basis of human social interaction: from concepts to brain imaging. *Physiol. Rev.* 89, 453–479. doi: 10.1152/physrev.00041.2007

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., and Keysers, C. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends Cogn. Sci.* 16, 114–121. doi: 10.1016/j.tics.2011.12.007

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science* 303, 1634–1640. doi: 10.1126/science.1089506

Hirata, M., Ikeda, T., Kikuchi, M., Kimura, T., Hiraishi, H., Yoshimura, Y., et al. (2014). Hyperscanning MEG for understanding mother-child cerebral interactions. *Front. Hum. Neurosci.* 8:118. doi: 10.3389/fnhum.2014.00118

Hirsch, J., Adam Noah, J., Zhang, X., Dravida, S., and Ono, Y. (2018). A cross-brain neural mechanism for human-to-human verbal communication. *Soc. Cogn. Affect. Neurosci.* 13, 907–920. doi: 10.1093/scan/nsy070

Hirsch, J., Zhang, X., Noah, J. A., and Ono, Y. (2017). Frontal temporal and parietal systems synchronize within and across brains during live eye-to-eye contact. *Neuroimage* 157, 314–330. doi: 10.1016/j.neuroimage.2017.06.018

Holper, L., Goldin, A. P., Shalóm, D. E., Battro, A. M., Wolf, M., and Sigman, M. (2013). The teaching and the learning brain: a cortical hemodynamic marker of teacher–student interactions in the socratic dialog. *Int. J. Educ. Res.* 59, 1–10. doi: 10.1016/j.ijer.2013.02.002

Holper, L., Scholkmann, F., and Wolf, M. (2012). Between-brain connectivity during imitation measured by fNIRS. *Neuroimage* 63, 212–222. doi: 10.1016/j.neuroimage.2012.06.028

Hu, Y., Hu, Y., Li, X., Pan, Y., and Cheng, X. (2017). Brain-to-brain synchronization across two persons predicts mutual prosociality. *Soc. Cogn. Affect. Neurosci.* 12, 1835–1844. doi: 10.1093/scan/nsx118

Hu, Y., Pan, Y., Shi, X., Cai, Q., Li, X., and Cheng, X. (2018). Inter-brain synchrony and cooperation context in interactive decision making. *Biol. Psychol.* 133, 54–62. doi: 10.1016/j.biopsycho.2017.12.005

Ikeda, S., Nozawa, T., Yokoyama, R., Miyazaki, A., Sasaki, Y., Sakaki, K., et al. (2017). Steady beat sound facilitates both coordinated group walking and inter-subject neural synchrony. *Front. Hum. Neurosci.* 11:147. doi: 10.3389/fnhum.2017.00147

Jiang, J., Chen, C., Dai, B., Shi, G., Ding, G., Liu, L., et al. (2015). Leader emergence through interpersonal neural synchronization. *Proc. Natl. Acad. Sci. U.S.A.* 112, 4274–4279. doi: 10.1073/pnas.1422930112

Jiang, J., Dai, B., Peng, D., Zhu, C., Liu, L., and Lu, C. (2012). Neural synchronization during face-to-face communication. *J. Neurosci.* 32, 16064–16069. doi: 10.1523/JNEUROSCI.2926-12.2012

Kawasaki, M., Yamada, Y., Ushiku, Y., Miyauchi, E., and Yamaguchi, Y. (2013). Inter-brain synchronization during coordination of speech rhythm in human-to-human social interaction. *Sci. Rep.* 3:1692. doi: 10.1038/srep01692

King-Casas, B. (2005). Getting to know you: reputation and trust in a two-person economic exchange. *Science* 308, 78–83. doi: 10.1126/science.1108062

Kinreich, S., Djalovski, A., Kraus, L., Louzoun, Y., and Feldman, R. (2017). Brain-to-brain synchrony during naturalistic social interactions. *Sci. Rep.* 7:17060. doi: 10.1038/s41598-017-17339-5

Koike, T., Sumiya, M., Nakagawa, E., Okazaki, S., and Sadato, N. (2019). What makes eye contact special? Neural substrates of on-line mutual eye-gaze: a hyperscanning fMRI study. *Eneuro* 6:ENEURO.0284–18.2019. doi: 10.1523/ENEURO.0284-18.2019

Koike, T., Tanabe, H. C., Okazaki, S., Nakagawa, E., Sasaki, A. T., Shimada, K., et al. (2016). Neural substrates of shared attention as social memory: a hyperscanning functional magnetic resonance imaging study. *Neuroimage* 125, 401–412. doi: 10.1016/j.neuroimage.2015.09.076

Koike, T., Tanabe, H. C., and Sadato, N. (2015). Hyperscanning neuroimaging technique to reveal the "two-in-one" system in social interactions. *Neurosci. Res.* 90, 25–32. doi: 10.1016/j.neures.2014.11.006

Konvalinka, I., Bauer, M., Stahlhut, C., Hansen, L. K., Roepstorff, A., and Frith, C. D. (2014). Frontal alpha oscillations distinguish leaders from followers: multivariate decoding of mutually interacting brains. *Neuroimage* 94, 79–88. doi: 10.1016/j.neuroimage.2014.03.003

Konvalinka, I., and Roepstorff, A. (2012). The two-brain approach: how can mutually interacting brains teach us something about social interaction? *Front. Hum. Neurosci.* 6:215. doi: 10.3389/fnhum.2012.00215

Lachat, F., Hugueville, L., Lemaréchal, J.-D., Conty, L., and George, N. (2012). Oscillatory brain correlates of live joint attention: a dual-EEG study. *Front. Hum. Neurosci.* 6:156. doi: 10.3389/fnhum.2012.00156

Levy, J., Goldstein, A., and Feldman, R. (2017). Perception of social synchrony induces mother-child gamma coupling in the social brain. *Soc. Cogn. Affect. Neurosci.* 12, 1036–1046. doi: 10.1093/scan/nsx032

Lindenberger, U., Li, S.-C., Gruber, W., and Müller, V. (2009). Brains swinging in concert: cortical phase synchronization while playing guitar. *BMC Neurosci.* 10:22. doi: 10.1186/1471-2202-10-22

Liu, N., Mok, C., Witt, E. E., Pradhan, A. H., Chen, J. E., and Reiss, A. L. (2016a). NIRS-based hyperscanning reveals inter-brain neural synchronization during cooperative Jenga game with face-to-face communication. *Front. Hum. Neurosci.* 10:82. doi: 10.3389/fnhum.2016.00082

Liu, T., Saito, H., and Oi, M. (2015). Role of the right inferior frontal gyrus in turn-based cooperation and competition: a near-infrared spectroscopy study. *Brain Cogn.* 99, 17–23. doi: 10.1016/j.bandc.2015.07.001

Liu, T., Saito, H., and Oi, M. (2016b). Obstruction increases activation in the right inferior frontal gyrus. *Soc. Neurosci.* 11, 344–352. doi: 10.1080/17470919.2015.1088469

Lopes, P. N., Salovey, P., Côté, S., and Beers, M. (2005). Emotion regulation abilities and the quality of social interaction. *Emotion* 5, 113–118. doi: 10.1037/1528-3542.5.1.113

Luck, S. J., and Hillyard, S. A. (1994). Spatial filtering during visual search: evidence from human electrophysiology. *J. Exp. Psychol. Hum. Percept. Perform.* 20:1000. doi: 10.1037/0096-1523.20.5.1000

Mandel, A., Bourguignon, M., Parkkonen, L., and Hari, R. (2016). Sensorimotor activation related to speaker vs. listener role during natural conversation. *Neurosci. Lett.* 614, 99–104. doi: 10.1016/j.neulet.2015.12.054

Melnik, A., Legkov, P., Izdebski, K., Kärcher, S. M., Hairston, W. D., Ferris, D. P., et al. (2017). Systems, subjects, sessions: to what extent do these factors influence EEG data? *Front. Hum. Neurosci.* 11:150. doi: 10.3389/fnhum.2017.00150

Ménoret, M., Varnet, L., Fargier, R., Cheylus, A., Curie, A., des Portes, V., et al. (2014). Neural correlates of non-verbal social interactions: a dual-EEG study. *Neuropsychologia* 55, 85–97. doi: 10.1016/j.neuropsychologia.2013.10.001

Michel, C. M., and Brunet, D. (2019). EEG source imaging: a practical review of the analysis steps. *Front. Neurol.* 10:325. doi: 10.3389/fneur.2019.00325

Montague, P. (2002). Hyperscanning: simultaneous fMRI during linked social interactions. *Neuroimage* 16, 1159–1164. doi: 10.1006/nimg.2002.1150

Mu, Y., Guo, C., and Han, S. (2016). Oxytocin enhances inter-brain synchrony during social coordination in male adults. *Soc. Cogn. Affect. Neurosci.* 11, 1882–1893. doi: 10.1093/scan/nsw106

Müller, V., and Lindenberger, U. (2014). Hyper-brain networks support romantic kissing in humans. *PLoS ONE* 9:e112080. doi: 10.1371/journal.pone.0112080

Müller, V., Sänger, J., and Lindenberger, U. (2013). Intra- and inter-brain synchronization during musical improvisation on the guitar. *PLoS ONE* 8:e73852. doi: 10.1371/journal.pone.0073852

Naeem, M., Prasad, G., Watson, D. R., and Kelso, J. S. (2012a). Electrophysiological signatures of intentional social coordination in the 10–12 Hz range. *Neuroimage* 59, 1795–1803. doi: 10.1016/j.neuroimage.2011.08.010

Naeem, M., Prasad, G., Watson, D. R., and Kelso, J. S. (2012b). Functional dissociation of brain rhythms in social coordination. *Clin. Neurophysiol.* 123, 1789–1797. doi: 10.1016/j.clinph.2012.02.065

Nguyen, T., Schleihauf, H., Kayhan, E., Matthes, D., Vrtička, P., and Hoehl, S. (2019). The effects of interaction quality on neural synchrony during mother-child problem solving. *Cortex* 124, 235–249. doi: 10.1016/j.cortex.2019.11.020

Novembre, G., Knoblich, G., Dunne, L., and Keller, P. E. (2017). Interpersonal synchrony enhanced through 20 Hz phase-coupled dual brain stimulation. *Soc. Cogn. Affect. Neurosci.* 12, 662–670. doi: 10.1093/scan/nsw172

Novembre, G., Sammler, D., and Keller, P. E. (2016). Neural alpha oscillations index the balance between self-other integration and segregation in real-time joint action. *Neuropsychologia* 89, 414–425. doi: 10.1016/j.neuropsychologia.2016.07.027

Nozawa, T., Sasaki, Y., Sakaki, K., Yokoyama, R., and Kawashima, R. (2016). Interpersonal frontopolar neural synchronization in group communication: an exploration toward fNIRS hyperscanning of natural interactions. *Neuroimage* 133, 484–497. doi: 10.1016/j.neuroimage.2016.03.059

Oliveira, A. S., Schlink, B. R., Hairston, W. D., König, P., and Ferris, D. P. (2016). Induction and separation of motion artifacts in EEG data using a mobile phantom head device. *J. Neural Eng.* 13:036014. doi: 10.1088/1741-2560/13/3/036014

Osaka, N., Minamoto, T., Yaoi, K., Azuma, M., and Osaka, M. (2014). Neural synchronization during cooperated humming: a hyperscanning study using fNIRS. *Proc. Soc. Behav. Sci.* 126, 241–243. doi: 10.1016/j.sbspro.2014.02.395

Osaka, N., Minamoto, T., Yaoi, K., Azuma, M., Shimada, Y. M., and Osaka, M. (2015). How two brains make one synchronized mind in the inferior frontal cortex: fNIRS-based hyperscanning during cooperative singing. *Front. Psychol.* 6:1811. doi: 10.3389/fpsyg.2015.01811

Pan, Y., Cheng, X., Zhang, Z., Li, X., and Hu, Y. (2017). Cooperation in lovers: an fNIRS-based hyperscanning study: cooperation in lovers. *Hum. Brain Mapp.* 38, 831–841. doi: 10.1002/hbm.23421

Pan, Y., Novembre, G., Song, B., Li, X., and Hu, Y. (2018). Interpersonal synchronization of inferior frontal cortices tracks social interactive learning of a song. *Neuroimage* 183, 280–290. doi: 10.1016/j.neuroimage.2018.08.005

Pérez, A., Carreiras, M., and Duñabeitia, J. A. (2017). Brain-to-brain entrainment: EEG interbrain synchronization while speaking and listening. *Sci. Rep.* 7:4190. doi: 10.1038/s41598-017-04464-4

Quaresima, V., and Ferrari, M. (2019). Functional near-infrared spectroscopy (fNIRS) for assessing cerebral cortex function during human behavior in natural/social situations: a concise review. *Organ. Res. Methods* 22, 46–68. doi: 10.1177/1094428116658959

Redcay, E., and Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nat. Rev. Neurosci.* 20, 495–505. doi: 10.1038/s41583-019-0179-4

Reindl, V., Gerloff, C., Scharke, W., and Konrad, K. (2018). Brain-to-brain synchrony in parent-child dyads and the relationship with emotion regulation revealed by fNIRS-based hyperscanning. *Neuroimage* 178, 493–502. doi: 10.1016/j.neuroimage.2018.05.060

Saito, D. N., Tanabe, H. C., Izuma, K., Hayashi, M. J., Morito, Y., Komeda, H., et al. (2010). "Stay tuned": inter-individual neural synchronization during mutual gaze and joint attention. *Front. Integr. Neurosci.* 4:127. doi: 10.3389/fnint.2010.00127

Sanfey, A. G. (2003). The neural basis of economic decision-making in the ultimatum game. *Science* 300, 1755–1758. doi: 10.1126/science.1082976

Sänger, J., Müller, V., and Lindenberger, U. (2012). Intra- and interbrain synchronization and network properties when playing guitar in duets. *Front. Hum. Neurosci.* 6:312. doi: 10.3389/fnhum.2012.00312

Sänger, J., Müller, V., and Lindenberger, U. (2013). Directionality in hyperbrain networks discriminates between leaders and followers in guitar duets. *Front. Hum. Neurosci.* 7:234. doi: 10.3389/fnhum.2013.00234

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., et al. (2013). Toward a second-person neuroscience. *Behav. Brain Sci.* 36, 393–414. doi: 10.1017/S0140525X12000660

Schippers, M. B., Roebroeck, A., Renken, R., Nanetti, L., and Keysers, C. (2010). Mapping the information flow from one brain to another during gestural communication. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9388–9393. doi: 10.1073/pnas.1001791107

Scholkmann, F., Holper, L., Wolf, U., and Wolf, M. (2013). A new methodical approach in neuroscience: assessing inter-personal brain coupling using functional near-infrared imaging (fNIRI) hyperscanning. *Front. Hum. Neurosci.* 7:813. doi: 10.3389/fnhum.2013.00813

Sebanz, N., Knoblich, G., and Prinz, W. (2003). Representing others' actions: just like one's own? *Cognition* 88, B11–B21. doi: 10.1016/S0010-0277(03)00043-X

Sebanz, N., Knoblich, G., Prinz, W., and Wascher, E. (2006). Twin peaks: an ERP study of action planning and control in coacting individuals. *J. Cogn. Neurosci.* 18, 859–870. doi: 10.1162/jocn.2006.18.5.859

Sinha, N., Maszczyk, T., Zhang Wanxuan, Tan, J., and Dauwels, J. (2016). "EEG hyperscanning study of inter-brain synchrony during cooperative and competitive interaction," in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (Budapest: IEEE), 004813–004818.

Spiegelhalder, K., Ohlendorf, S., Regen, W., Feige, B., Tebartz van Elst, L., Weiller, C., et al. (2014). Interindividual synchronization of brain activity during live verbal communication. *Behav. Brain Res.* 258, 75–79. doi: 10.1016/j.bbr.2013.10.015

Sporns, O. (2018). Graph theory methods: applications in brain networks. *Dialog. Clin. Neurosci.* 20:111–121.

Stephens, G. J., Silbert, L. J., and Hasson, U. (2010). Speaker–listener neural coupling underlies successful communication. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14425–14430. doi: 10.1073/pnas.1008662107

Szymanski, C., Müller, V., Brick, T. R., von Oertzen, T., and Lindenberger, U. (2017a). Hyper-transcranial alternating current stimulation: experimental manipulation of inter-brain synchrony. *Front. Hum. Neurosci.* 11:539. doi: 10.3389/fnhum.2017.00539

Szymanski, C., Pesquita, A., Brennan, A. A., Perdikis, D., Enns, J. T., Brick, T. R., et al. (2017b). Teams on the same wavelength perform better: inter-brain phase synchronization constitutes a neural substrate for social facilitation. *Neuroimage* 152, 425–436. doi: 10.1016/j.neuroimage.2017.03.013

Tanabe, H. C., Kosaka, H., Saito, D. N., Koike, T., Hayashi, M. J., Izuma, K., et al. (2012). Hard to "tune in": neural mechanisms of live face-to-face interaction with high-functioning autistic spectrum disorder. *Front. Hum. Neurosci.* 6:268. doi: 10.3389/fnhum.2012.00268

Tang, H., Mai, X., Wang, S., Zhu, C., Krueger, F., and Liu, C. (2016). Interpersonal brain synchronization in the right temporo-parietal junction during face-to-face economic exchange. *Soc. Cogn. Affect. Neurosci.* 11, 23–32. doi: 10.1093/scan/nsv092

Thatcher, R. W. (2012). Coherence, phase differences, phase shift, and phase lock in eeg/erp analyses. *Dev. Neuropsychol.* 37, 476–496. doi: 10.1080/87565641.2011.619241

Tognoli, E., Lagarde, J., DeGuzman, G. C., and Kelso, J. A. S. (2007). The phi complex as a neuromarker of human social coordination. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8190–8195. doi: 10.1073/pnas.0611453104

Tomlin, D. (2006). Agent-specific responses in the cingulate cortex during economic exchanges. *Science* 312, 1047–1050. doi: 10.1126/science.1125596

Tomlin, D., Nedic, A., Prentice, D. A., Holmes, P., and Cohen, J. D. (2013). The neural substrates of social influence on decision making. *PLoS ONE* 8:e52630. doi: 10.1371/journal.pone.0052630

Twenge, J. M., Baumeister, R. F., DeWall, C. N., Ciarocco, N. J., and Bartels, J. M. (2007). Social exclusion decreases prosocial behavior. *J. Pers. Soc. Psychol.* 92, 56–66. doi: 10.1037/0022-3514.92.1.56

Wahn, B., Czeszumski, A., and König, P. (2018). Performance similarities predict collective benefits in dyadic and triadic joint visual search. *PLos ONE* 13:e0191179. doi: 10.1371/journal.pone.0191179

Wahn, B., Czeszumski, A., Labusch, M., Kingstone, A., and König, P. (2019). Dyadic and triadic search: benefits, costs, and predictors of group performance. *PsyArXiv*. doi: 10.31234/osf.io/894zt

Wang, C., Zhang, T., Shan, Z., Liu, J., Yuan, D., and Li, X. (2019). Dynamic interpersonal neural synchronization underlying pain-induced cooperation in females. *Hum. Brain Mapp.* 40, 3222–3232. doi: 10.1002/hbm.24592

Wen, X., Mo, J., and Ding, M. (2012). Exploring resting-state functional connectivity with total interdependence. *Neuroimage*. 60, 1587–1595. doi: 10.1016/j.neuroimage.2012.01.079

Yun, K., Chung, D., and Jeong, J. (2008). "Emotional interactions in human decision making using EEG hyperscanning," In *International Conference of Cognitive Science* (Seoul), 4.

Yun, K., Watanabe, K., and Shimojo, S. (2012). Interpersonal body and neural synchronization as a marker of implicit social interaction. *Sci. Rep.* 2:959. doi: 10.1038/srep00959

Zamm, A., Debener, S., Bauer, A.-K. R., Bleichner, M. G., Demos, A. P., and Palmer, C. (2018). Amplitude envelope correlations measure synchronous cortical oscillations in performing musicians: amplitude envelopes measure inter-brain synchrony. *Ann. N. Y. Acad. Sci.* 1423, 251–263. doi: 10.1111/nyas.13738

Zhang, M., Liu, T., Pelowski, M., and Yu, D. (2017). Gender difference in spontaneous deception: a hyperscanning study using functional near-infrared spectroscopy. *Sci. Rep.* 7:7508. doi: 10.1038/s41598-017-06764-1

Zhdanov, A., Nurminen, J., Baess, P., Hirvenkari, L., Jousmäki, V., Mäkelä, J. P., et al. (2015). An internet-based real-time audiovisual link for dual MEG recordings. *PLoS ONE* 10:e0128485. doi: 10.1371/journal.pone.0128485

Zhou, G., Bourguignon, M., Parkkonen, L., and Hari, R. (2016). Neural signatures of hand kinematics in leaders vs. followers: a dual-MEG study. *Neuroimage* 125, 731–738. doi: 10.1016/j.neuroimage.2015.11.002

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Human Neuroscience

ORIGINAL RESEARCH
published: 03 March 2020
doi: 10.3389/fnhum.2020.00057

Developmental Differences in Cortical Activation During Action Observation, Action Execution and Interpersonal Synchrony: An fNIRS Study

*Wan-Chun Su[1,2], McKenzie L. Culotta[1,2], Michael D. Hoffman[1], Susanna L. Trost[1], Kevin A. Pelphrey[3], Daisuke Tsuzuki[4] and Anjana N. Bhat[1,2,5]**

[1]Department of Physical Therapy, University of Delaware, Newark, DE, United States, [2]Biomechanics & Movement Science Program, University of Delaware, Newark, DE, United States, [3]Department of Neurology & The UVA Brain Institute, University of Virginia, Charlottesville, VA, United States, [4]Department of Language Sciences, Tokyo Metropolitan University, Tokyo, Japan, [5]Behavioral Neuroscience Program, Department of Psychological & Brain Sciences, University of Delaware, Newark, DE, United States

Edited by:
Peter König,
University of Osnabrück, Germany

Reviewed by:
Muhammad Jawad Khan,
National University of Sciences and Technology, Pakistan
Keum-Shik Hong,
Pusan National University, South Korea
Rick Dale,
University of California, Los Angeles, United States

***Correspondence:**
Anjana N. Bhat
abhat@udel.edu

Specialty section:
This article was submitted to Brain Imaging and Stimulation, a section of the journal Frontiers in Human Neuroscience

Received: *22 May 2019*
Accepted: *06 February 2020*
Published: *03 March 2020*

Citation:
Su W-C, Culotta ML, Hoffman MD, Trost SL, Pelphrey KA, Tsuzuki D and Bhat AN (2020) Developmental Differences in Cortical Activation During Action Observation, Action Execution and Interpersonal Synchrony: An fNIRS Study. Front. Hum. Neurosci. 14:57. doi: 10.3389/fnhum.2020.00057

Interpersonal synchrony (IPS) is an important everyday behavior influencing social cognitive development; however, few studies have investigated the developmental differences and underlying neural mechanisms of IPS. functional near-infrared spectroscopy (fNIRS) is a novel neuroimaging tool that allows the study of cortical activation in the presence of natural movements. Using fNIRS, we compared cortical activation patterns between children and adults during action observation, execution, and IPS. Seventeen school-age children and 15 adults completed a reach to cleanup task while we obtained cortical activation data from bilateral inferior frontal gyrus (IFG), superior temporal sulcus (STS), and inferior parietal lobes (IPL). Children showed lower spatial and temporal accuracy during IPS compared to adults (i.e., spatial synchrony scores (Mean ± SE) in children: 2.67 ± 0.08 and adults: 2.85 ± 0.06; temporal synchrony scores (Mean ± SE) in children: 2.74 ± 0.06 and adults: 2.88 ± 0.05). For both groups, the STS regions were more activated during action observation, while the IFG and STS were more activated during action execution and IPS. The IPS condition involved more right-sided activation compared to action execution suggesting that IPS is a higher-order process involving more bilateral cortical activation. In addition, adults showed more left lateralization compared to the children during movement conditions (execution and IPS); which indicated greater inhibition of ipsilateral cortices in the adults compared to children. These findings provide a neuroimaging framework to study imitation and IPS impairments in special populations such as children with Autism Spectrum Disorder.

Keywords: development, fNIRS, imitation, interpersonal synchrony, lateralization

INTRODUCTION

Interpersonal synchrony (IPS) or the time-constrained movement coordination between two individuals is an important daily activity (Repp and Su, 2013). Some examples of IPS in daily life include walking or running at a matching pace with a partner, two people lifting a large or heavy object together, and children playing "follow the leader" games. Musicians often synchronize their actions while playing instruments in order to achieve harmony (Phillips-Silver and Keller, 2012). IPS has been studied across a variety of tasks spanning from simple finger tapping and reaching for objects (Rabinowitch and Knafo-Noam, 2015; Schmitz et al., 2017) to whole-body swaying/rocking (Sofianidis et al., 2012; Marsh et al., 2013) as well as walking (Wiltermuth and Heath, 2009). Yet, few studies have examined the developmental differences in IPS between children and adults. Moreover, the underlying neural mechanisms of IPS have not been well studied. In this study, we compared IPS performance and associated cortical activation patterns using functional near-infrared spectroscopy (fNIRS) between typically developing (TD) young adults and school-age children.

In the first 2 years of life, infants transition from imitation of discrete actions that are one-step and familiar to those that are multi-step and unfamiliar in nature (Jones, 2007). By 2 years, toddlers perform various sustained rhythmic actions such as walking, running, drumming, etcetera (Clark and Phillips, 1993; Brakke et al., 2007). Preschoolers as young as two and a half years of age were able to scale their drumming tempo to that of their social partner (Kirschner and Tomasello, 2009). In a different study, young elementary school child-child pairs showed the lowest levels of IPS during joint drumming followed by middle school child-child pairs and lastly the young adult-adult pairs (Kleinspehn-Ammerlahn et al., 2011). The lower IPS levels of young children were attributed to their difficulties in adjusting to the variable nature of their partner's hand coordination patterns (Kleinspehn-Ammerlahn et al., 2011). Infants and children will learn a variety of important social and adaptive skills by engaging in imitation and IPS with their social partners (Carpenter et al., 1998; Meltzoff, 2007). While short bouts of synchronization contribute to greater social bonding and pro-social behavior (Macrae et al., 2008; Tunçgenç and Cohen, 2016a,b), long-term exchanges of parent-child synchrony experiences will help develop secure attachments with caregivers (Isabella and Belsky, 1991). Children who engaged in more synchronous clap-tap actions had more prosocial behaviors towards their peers than those who had less synchronous actions (Tunçgenç and Cohen, 2016b). A broader meta-analysis of effects of IPS reported a medium-size effect on prosocial behaviors, a small-to-medium size effect on social bonding such as a greater sense of affiliation/similarity as well as better social cognition, for example, better memory of the partner (Mogan et al., 2017). By comparing the IPS performance and associated cortical activation during a novel and continuous reach to cleanup task between children and adults, the present study will highlight the developmental differences in IPS and related neural mechanisms.

While there are few studies describing neural substrates underlying IPS behaviors, various cortical structures have been implicated in imitation behaviors, and both behaviors may share similar neural substrates for their control (Bhat et al., 2017). Various cortical regions play an important role during the process of imitation (Iacoboni, 2005; Cattaneo and Rizzolatti, 2009). These include the frontal regions of the Inferior Frontal Gyrus (IFG) and ventral Premotor Cortex, the parietal regions such as the Inferior Parietal Lobule (IPL) and intraparietal sulcus of the parietal lobe, specifically, the Inferior Parietal Gyrus, Supramarginal Gyrus, and Angular Gyrus, and the temporal regions, specifically, the Superior Temporal Sulcus (STS; Iacoboni, 2005; Cattaneo and Rizzolatti, 2009). The STS responds more to biological than nonbiological stimuli and is thought to encode biological motion rather than the superficial characteristics of moving stimuli (Pelphrey et al., 2003a). Greater bilateral STS activation was found during imitation than action observation and execution suggesting that it provides a visual description of observed actions and compares the observed movements to that of planned actions (Molenberghs et al., 2010). On the other hand, IFG and IPL regions are said to be more active during the observation and imitation of goal-directed, object-related actions (Iacoboni, 2005; Pokorny et al., 2015). IPL may contribute to the motor planning aspects of imitated actions (De Renzi et al., 1983; Fontana et al., 2012); while IFG is said to be responsible for processing the goals of the action (Koski et al., 2002). It is important to note that during imitation these regions do not work in isolation and instead interact with each other as well as other brain regions such as the dorsolateral prefrontal cortices, premotor cortices, primary and supplementary/pre-supplementary motor cortices, cingulate/insular cortices, cuneus/precuneus as well as subcortical structures such as the cerebellum and putamen to form an important imitation network (Gazzola and Keysers, 2009; Iacoboni, 2009). Given the important role of the aforementioned cortical regions for imitation performance, in the present study, we will assess their contributions to components of IPS behaviors including action observation, action execution, and IPS itself during a continuous reach to cleanup task in both adults and children.

Original studies comparing cortical activation during action observation, execution, and imitation have reported similar levels of activation across all three tasks (Cattaneo and Rizzolatti, 2009; Molenberghs et al., 2010). However, recent studies have reported a more variable level of activation and lateralization across observation, execution, and imitation tasks. In terms of level of activation, one study reported greatest cortical activation during action imitation followed by action execution and lowest activation during action observation (Aziz-Zadeh et al., 2006) while another study found that cortical activation was greater during action execution and imitation than action observation (Reynolds et al., 2019). Multiple studies have also reported greater cortical activation during action imitation than action execution and observation (Hamzei et al., 2016; Brihmat et al., 2017; Gatti et al., 2017). However, the aforementioned patterns of activation differ depending on the regions of interest (ROIs). For example, Montgomery et al. (2007) found that IPL and IFG were more active during action imitation and execution compared to action observation whereas STS activation was greater during action imitation compared to action observation and execution. Varying patterns of hemispheric lateralization have also been reported for imitation behaviors. One of the original studies by Aziz-Zadeh et al. (2006) suggested that imitation control is more bilateral in nature. Other studies had shown that during action imitation STS activated bilaterally whereas IFG and IPL activation was more variable depending on the nature of the task (Mühlau et al., 2005; Montgomery et al., 2007; Gatti et al., 2017). During imitation of gesture or goal-directed actions, Mühlau et al. (2005) and Montgomery et al. (2007) found greater activation over left than right IPL but similar IFG activation between hemispheres, while Gatti et al. (2017) found greater activation in the right precentral gyrus and right IFG compared to their left homologues. In spite of the variable findings of past fMRI studies, they do offer some evidence for how the different cortical regions play a role during imitation and this could perhaps extend to IPS behaviors as well. However, the fMRI environment limits the study tasks to simple hand gestures without face-to-face social interactions. We still do not know if the aforementioned findings can be generalized to complex, everyday motor tasks within naturalistic social contexts.

fNIRS is a fairly novel neuroimaging technique that measures cerebral hemodynamics similar to fMRI, the gold-standard of neuroimaging (Lloyd-Fox et al., 2010). But unlike fMRI, which requires the participant to lie still in a narrow scanning bore, fNIRS only restrains a participant through a cap on the head and allows for measurements in the presence of movement as well as face-to-face interactions. Given its advantage to tolerate motion artifacts, fNIRS has been used to study cortical activation across various movements of walking (Holtzer et al., 2019), playing a dance video game (Tachibana et al., 2011), as well as free arm movements during face to face interactions with others (Egetemeir et al., 2011). Moreover, with its greater temporal resolution compared to fMRI (Lloyd-Fox et al., 2010), fNIRS does a better job of detecting the onset and features of the hemodynamic response (Hong et al., 2018; Khan et al., 2018). This ability to distinguish features of the hemodynamic response may help in identifying differences related to development as well as neuropathology. Other sophisticated applications of fNIRS include the use of fNIRS-based hemodynamic responses to facilitate human-computer interaction (Naseer and Hong, 2015) as well as multimodal use of fNIRS and EEG (Ge et al., 2017, 2019). Specifically, greater STS activation, as well as the larger amplitude of EEG-based evoked response potentials, have been reported during observation of intentional grasping compared to meaningless grasping (Ge et al., 2019). A handful of studies have examined cortical activation during naturalistic face-to-face IPS and social cooperation/competition using fNIRS (Egetemeir et al., 2011; Bolling et al., 2013; Liu et al., 2015; Bhat et al., 2017). During a cooperation/competition game, the cooperators showed greater right IFG activation compared to the competitors (Liu et al., 2015). Similarly, adults showed greater IPL activation during joint action with a partner during a table-setting task compared to solo table-setting motions (Egetemeir et al., 2011). We have replicated the work of Egetemeir et al. (2011) in healthy adults by comparing lateral cortical activation during observation, execution, and synchronization of a reach and clean up task (Bhat et al., 2017). We too found greater activation in cortical regions of STS, IFG, and IPL during action execution and IPS compared to action observation. More importantly, right IFG and IPL regions were more active during IPS than the action execution condition. We concluded that the action execution condition led to more left-lateralized cortical activation whereas the IPS condition led to more bilateral cortical activation suggesting an important role for the right frontoparietal networks during IPS behaviors.

Considering the important role of IPS in facilitating social development, it would be valuable to study developmental differences in IPS and associated patterns of activation. To date, few studies have compared IPS behaviors during naturalistic reaching tasks as well as the underlying brain activation patterns between TD adults and TD children to describe the developmental differences in IPS. Therefore, in this study, we aimed to investigate the differences in brain activation between TD adults and TD children as they observed, executed, and synchronized actions during a reach-cleanup task. We hypothesized that the quality of IPS in TD children would differ from TD adults. Specifically, we expected the level of IPS to be lower and patterns of cortical activation to somewhat differ between adults and children. However, we expect both groups to have greater bilateral activation during the IPS condition compared to action execution. Thirdly, we also expected synchrony performance to correlate with cortical activation.

MATERIALS AND METHODS

Participants

Seventeen TD school-age children (mean age ± SE: 10.82 ± 0.69, 11 males and six females) and 15 TD adults participated in this study (mean age ± SE: 22.6 ± 0.7, eight males and seven females, $p < 0.001$ for the age difference between groups, no gender-based differences between groups, $p > 0.1$, **Table 1**). Individuals were recruited through word of mouth, online postings in local listservs as well as fliers in the community. As a first step, we completed screening interviews with potential participants to exclude individuals with any known neurological or psychiatric diagnoses, or those taking psychotropic medications, or any other difficulties that would prevent them from performing the study tasks. All participants had normal or corrected to normal vision. Based on a standard handedness questionnaire (Coren, 1992), 15 of the child participants were found to be strongly right-handed, while two children were moderately left-handed. Fourteen adult participants were strongly right-handed with one adult being weakly right-handed (**Table 1**). The activation patterns of the two moderately left-handed children and the weakly right-handed adult were similar to the group results as all had consistently used their right hand for completing the task; hence, their data have been retained following data analysis.

All participants completed the Vineland Adaptive Behavioral Scales (Volkmar et al., 1987) to provide measures of socialization (averaged standard score ± SE: children = 106.53 ± 3.18; adult = 106.53 ± 2.05, group difference: $p > 0.1$), communication

TABLE 1 | Demographic and developmental/cognitive data.

Characteristics	Child (n = 17) Mean ± SE	Adult (n = 15) Mean ± SE
Age	10.82 ± 0.69*	22.60 ± 0.70
Gender	11 male, 6 female	8 male, 7 female
Ethnicity	13 C, 1 A, 1 AI, 2 AC	12 C, 2 A, 1 Af
Handedness	15 R, 2 L	14 R, 1 L
VABS-II (SS)	110.29 ± 2.92	111.07 ± 2.53
Communication (SS)	109.82 ± 2.88	105.47 ± 1.65
Daily living (SS)	110.41 ± 3.08	110.07 ± 2.31
Socialization (SS)	106.53 ± 3.18	106.53 ± 2.05

*VABS-II, Vineland Adaptive Behavior Scale-2nd Edition; SS, standard score; SE, Standard error; M, Male, F, Female; C, Caucasian, A, Asian; AI, American Indian; AC, Asian-Caucasian; Af, African American; R, right; L, left. *Indicates a significant difference between groups.*

(children = 109.82 ± 2.88; adult = 105.47 ± 1.65, group difference: $p > 0.1$), daily living skills (children = 110.41 ± 3.08; adult = 110.07 ± 2.31, group difference: $p > 0.1$) as well as overall adaptive functioning (children = 110.29 ± 2.92; adult = 111.07 ± 2.53, group difference: $p > 0.1$). Both groups showed typical levels of subdomain and overall adaptive functioning with no significant differences between groups (**Table 1**). The University of Delaware Institutional Review Board (IRB) approved this study protocol. Procedures were carried out in accordance with the recommendations of our IRB (IRB protocol id #: 1227966-1). All adult participants gave written informed consent, the parents of child participants approved their child's participation, and the children gave their written assent as well, in accordance with the Declaration of Helsinki (as of 2008), prior to participation.

Experimental Procedures

Each participant and tester sat at a table facing each other to complete a reach to the cleanup task using a randomized blocked design (Bhat et al., 2017). Two 3 × 3 probes embedded in a cap were placed on the participant's head (**Figures 1A,B**). Eight colored blocks were placed on a mat in a circular manner in front of both, the participant and the tester. Participants were asked to clean up the blocks off the mat into a bowl placed on the right using their right hand only. The participant completed three conditions: WATCH, DO, and TOGETHER (**Figures 1A–C**). During the WATCH condition, the participant observed the tester pick up the blocks in a sequential manner and put them into the container. Adults generally paid attention to the task; however, to ensure that the children paid attention during the WATCH trials, we asked them to focus on the pattern of cleanup. Before the trial, children were asked to pay attention to how the cleanup was performed. After a WATCH trial was completed, they were asked, "Which block did I pick up first?" Or "which block did I pick up last?" Or "how did I clean up the blocks?" For the DO condition, the participants cleaned up all the blocks in a sequence of their choice. In the TOGETHER condition, the tester led the block cleanup in random order while the participant followed by picking up the same block as the tester. No questions were asked after completing the DO and TOGETHER conditions. The participant was asked to use their right hand; while the tester used their left hand. The adults completed a total of 24 trials (eight trials per condition) whereas the children completed a total of 18 trials (six trials per condition). The stimulation period ranged between 10 and 13 s [Duration in seconds (Mean ± SE) in adults: W = 11.5 ± 0.18; D = 11.2 ± 0.3; T = 13.8 ± 0.6 and duration in children: W = 10.6 ± 0.2; D = 10.3 ± 0.4; T = 13.6 ± 0.6; $p > 0.1$ for group differences]. A 10-s pre-stimulation and a 16-s post-stimulation period were included to account for any baseline drifts in the fNIRS signal and to allow the hemodynamic response to return to baseline before starting the next trial. During baseline periods, the participants were asked to focus on a cross-hair on the front wall and remain as still as possible.

Data Collection

The hemodynamic changes over the ROIs were recorded using the Hitachi ETG-4000 system (Hitachi Medical Systems, Tokyo, Japan), with a sampling rate of 10 Hz. Two 3 × 3 probe sets, consisting of five infrared emitters and four receivers (i.e., 24 channels), were positioned over bilateral frontoparietal and temporal regions. Each adjacent pair of probes that were 3 cm apart were an emitter and receiver of two wavelengths of infrared light (695 and 830 nm). The middle column of the probe set was aligned with the tragus of the ear and the lowermost row of the optode set was aligned with the T3 position of the International 10-20 system (Klem et al., 1999; **Figures 2A,B**).

The infrared light passed through the skull creating a banana-shaped arc and reached the cortical area approximately below the midpoint of the two probes. The attenuation of infrared light was used to calculate the changes in concentrations of oxygenated (HbO_2) and deoxygenated hemoglobin (HHb) chromophores per channel using the Modified Beer-Lambert Law. Based on results from previous studies, an increase in HbO_2 concentration and a decrease in HHb concentration were expected with increased brain activation within a certain ROI (Lloyd-Fox et al., 2010).

E-Prime presentation software (version 2.0) was used to trigger the Hitachi fNIRS system. The entire session was videotaped using a camcorder that was synchronized with the Hitachi fNIRS system.

Spatial Registration Approach

During the 3D registration process, each child was asked to remain in a still and upright position. The 3D locations of the standard cranial landmarks (nasion, inion, left and right preauricular points, and the Cz position of the International 10-20 system) as well as 3D locations of each probe in the probe set were recorded w.r.t. a reference coordinate system using the ETG-4000 3D positioning unit. These 3D coordinates saved in a text file format for each participant were run through MATLAB codes developed by the sixth author. The anchor-based, spatial registration method developed by Tsuzuki et al. (2012) was used to transform the 3D spatial location of each channel from the reference coordinate system to the Montreal Neurological Institute (MNI)'s coordinate system (see **Figure 2**, **Supplementary Table S1**). Structural information from an anatomical database (Okamoto et al., 2004) was used to

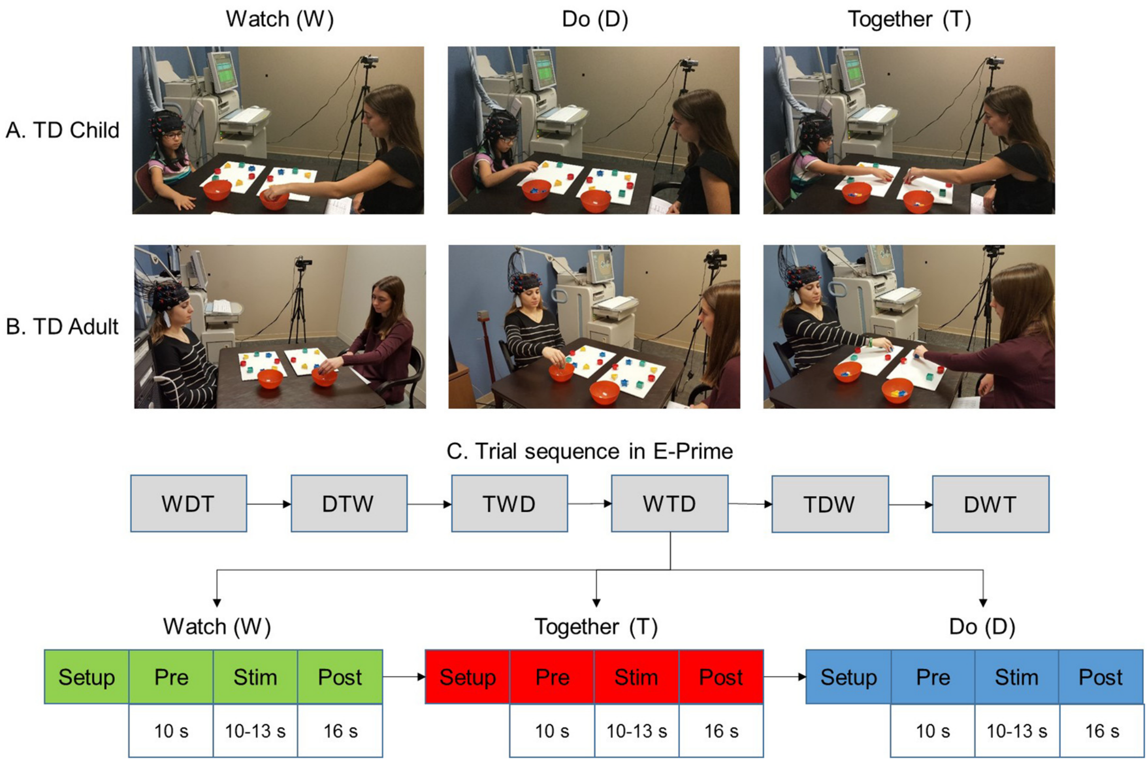


FIGURE 1 | Experimental setup **(A,B)** and task sequence **(C)**. Written permission for publication of participant pictures has been taken.

provide estimates of channel positions within a standardized 3D brain atlas (Tsuzuki et al., 2012). The estimated channel locations were anatomically labeled using the LONI Probabilistic Brain Atlas (LPBA; Shattuck et al., 2008). Note that each run includes position data from all participants within one group to obtain the average MNI coordinates for each channel. Based on the regions covered by our channels, we assigned the 24 channels to three ROIs for the children. Similarly, please refer to Bhat et al. (2017) for the channel assignments in adults across the same three ROIs.

The three ROIs included: (i) the Superior Anterior region (SA) which included channels over the inferior/middle frontal gyrus or IFG and pre-central gyrus (or frontal cortices, left: channels 1 and 3, and right: channels 14 and 17 channels, see **Figures 2C,D**); (ii) the Inferior Posterior region (IP) which included channels over the post-central gyrus, supramarginal gyrus, and angular gyrus (or the inferior posterior parietal cortices or IPL, left: 2, 4, 5, 7 channels and right: 13, 15, 16, 18 channels, see **Figures 2C,D**); and (iii) the Inferior Anterior region (IA) which included channels over the middle and superior temporal gyrus (or superior temporal cortices or STS, left: 10, 11, 12 channels and right: 20, 23, 24 channels, see **Figures 2C,D**). These three ROIs separated the three cortical regions we described earlier. Channels 6, 8, 9 (left) and 19, 21, and 22 (right) were excluded due to spatial uncertainty. To be clear, spatial uncertainty occurred when either one of the two homologous channels did not fall within the same ROI for a particular group. Another reason for spatial uncertainty was when any given channel did not cover 60% or more of the assigned ROI and instead covered multiple ROIs evenly, for example, 50% IPL and 50% IFG; such channels were excluded. In this way, we were able to consistently assign 18 out of the 24 channels to one of the aforementioned ROIs in both groups. Note: **Supplementary Table S1** in the "**Supplementary Materials**" section shows the channel assignment in children and refers to Bhat et al. (2017) for channel assignment in adults.

Data Processing

We have written our own customized MATLAB programs that incorporate functions from open-source software such as Hitachi POTATo (Sutoko et al., 2016) and Homer-2 (Huppert et al., 2009) to analyze the data from the ETG system (see data processing steps in **Figure 3**). The sampling frequency of the fNIRS system was 10 Hz (i.e., 10 data frames per second were collected). Data from each channel was first band-pass filtered between 0.01 and 0.5 Hz using the Fast Fourier Transform (FFT) method to remove lower or higher frequencies associated with body movements and other dynamic signals/tissue such as respiration, heart rate, skin blood flow, etc. The low-pass filter removes physiological noises related to respiration and fast cardiac oscillations and high-frequency instrument noise, whereas the high-pass filter minimizes the low-frequency drift from the data. To remove motion artifacts, we used the wavelet method (Sato et al., 2006; Huppert et al., 2009) which is one of

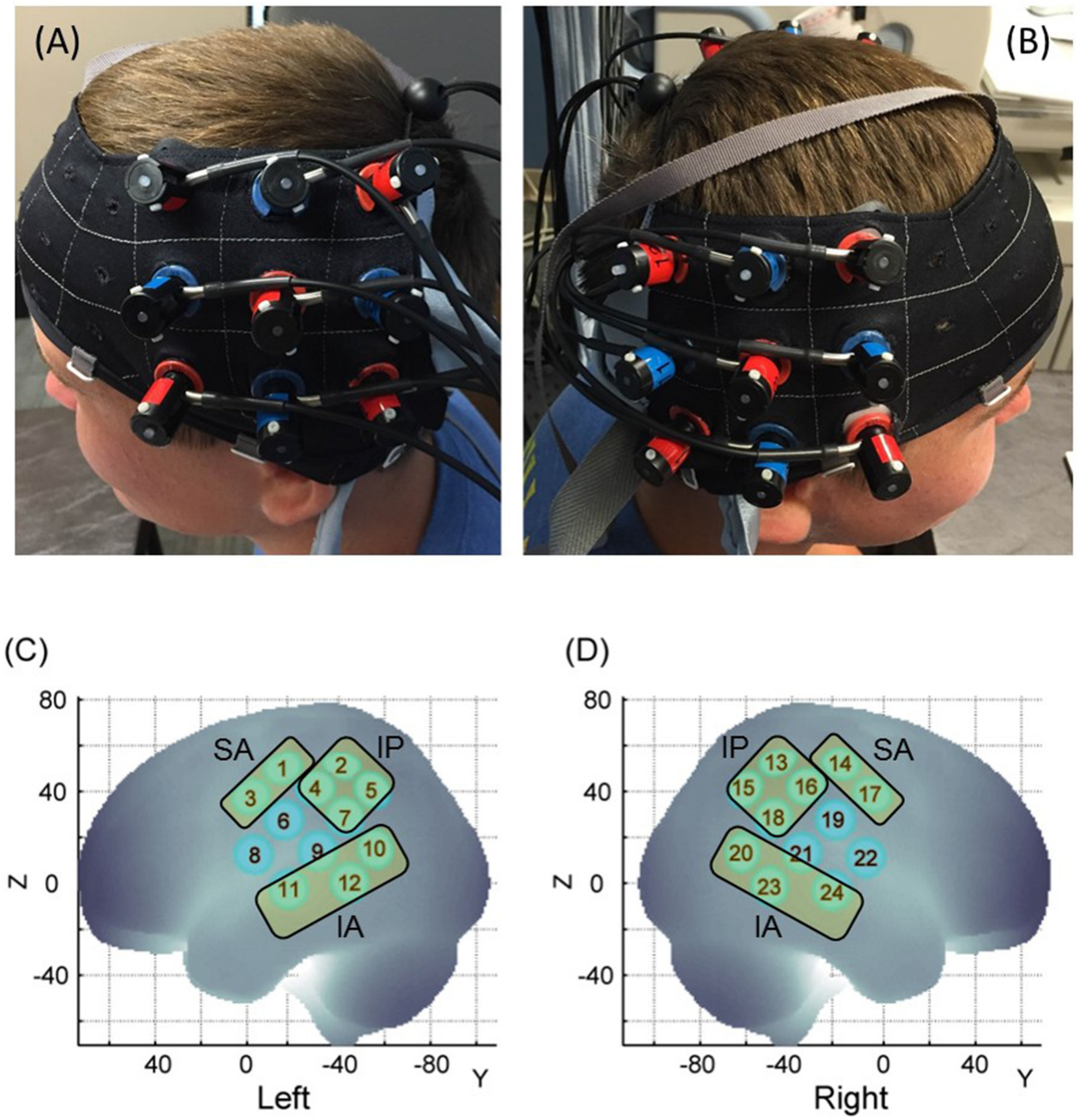


FIGURE 2 | Probe placement **(A,B)** and spatial registration output **(C,D)**. Written permission for publication of participant pictures has been taken.

the most robust methods for motion artifact removal (Hu et al., 2015). In this method, it is assumed that the measured signal is a linear combination of the desired signal and the undesired artifacts. The number of levels for wavelet decomposition is calculated by taking the logarithm of the number of data points using a base of 2. For our dataset, this value was 14. By applying the 1-D discrete wavelet transform to the signal from each channel, details of the signal are estimated as approximation coefficients. Assuming that the detail wavelet coefficients have a Gaussian distribution, outliers in the distribution correspond to the coefficients related to the motion artifacts. To identify the motion artifacts/outliers, an "iqr" parameter of 1.5 was used. The coefficients greater than the iqr parameter times the interquartile range of the data are typically associated with motion artifacts, and hence, they were set to zero to remove such artifacts. The inverse discrete wavelet transform is applied and the signal is reconstructed. Next, the General Linear Model (GLM) was implemented using a HOMER-2 MATLAB function. GLM estimated the hemodynamic response function using Gaussian basis functions and a 3$^{\text{rd}}$ order polynomial drift regression (Huppert et al., 2009). To correct the baseline drifts, the linear trend between the pre-trial baseline and the post-trial baseline was calculated and subtracted from values in the stimulation period as implemented in Hitachi POTATo (Sutoko et al., 2016). Average HbO_2 and HHb values were obtained for the stimulation period of each trial. The range of HbO_2 data was significantly

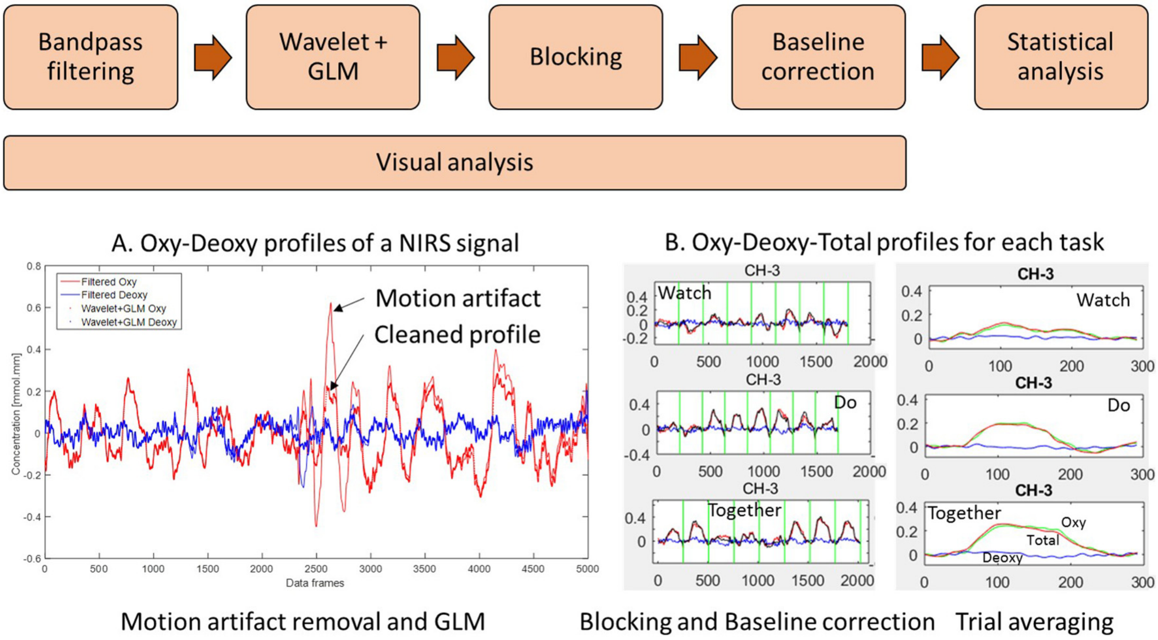


FIGURE 3 | Data processing workflow: **(A)** filter, wavelet and general linear model (GLM) of NIRS signal and **(B)** trial-by-trial view and Average view of Oxy Hb (HbO_2), Deoxy Hb (HHb), and Total Hb (HbT) profiles for a given channel. (W, D, T) from 5 s before to 24 s after the start of stimulation. Data have been averaged across trials and participants.

greater than the HHb data. Moreover, HbO_2 profiles have a greater signal to noise ratio compared to HHb and therefore consistent with fNIRS literature, we have reported HbO_2 profiles (Sato et al., 2006). The data were plotted and saved at each step. We visually screened the plotted figures at each step of the analysis to exclude channels/trials. We excluded channels with poor contact (flat lines) or persistent motion artifacts or obvious outliers compared to the other similar trials from each condition. Nearly 6.7% of data from children and 19.2% of data from adults have been removed using these criteria. In the "**Supplementary Materials**" section, we have also provided a visual representation of the second-to-second HbO_2 profile for each group (**Supplementary Figure S1**: Adults, **Supplementary Figure S2**: Children), each condition, and each channel for the entire period (pre-baseline, stimulation, and post-baseline). The pink vertical line denotes the start of the stimulation period and the data shown to the right of the pink line are the 240 frames across stimulation (10–13 s) and post-stimulation baseline (14–11 s) periods.

Video Data Coding

Two trained student researchers scored the session videos in order to exclude trials with significant errors. Inter-rater reliability of above 85% and intra-rater reliability of above 90% were achieved for all scores between the primary and secondary coder for 20% of the video dataset. The trials were excluded from data analysis if the participants did not follow task instructions, had significant body movements unrelated to the task, or spoke to the tester during the trials. A three-point scale was used to code the synchrony and motor quality during IPS. Spatial synchrony scores were rated from 1 to 3 with 1 = Picked up incorrect blocks for more than once, 2 = Picked up the incorrect block once, and 3 = Picked up all blocks correctly. Temporal synchrony scores were rated from 1 to 3 with 1 = More than one block delay, 2 = One-block delay and 3 = Perfect synchrony. Motor errors were defined as two-hand use, picking up more than one block at the same time, slippage when picking or placing, while the motor scores were rated with 1 = More than 4 errors, 2 = 2–4 errors, 3 = 0–1 error. The number of additional movements during the stimulation period was also coded. Ultimately, we eliminated 7.0% of the overall child data and 1.1% of the overall adult data due to persistent motion artifacts. Specifically, in the TD child group, 7.3% of Watch, 9.8% of Do, and 3.9% of Together or 1–2 trials out of the six trials for each condition were excluded. In the TD adult group, 0% of Watch, 0% of Do, and 3.3% of Together or approximately 1–2 out of the eight trials per condition were excluded.

Statistical Analyses

To avoid multiple channel-wise comparisons, we averaged data across channels within the same ROI based on our spatial registration output (**Figures 2C,D** show the six ROIs and constituent channels). All participants primarily moved their right hand during the task, therefore, right hemisphere activation should be considered ipsilateral, and left hemisphere activation would be contralateral to the moving arm of our participants. We determined levels of activation for six ROIs including the left and right superoanterior (SA), inferoposterior (IP), and inferoanterior (IA) regions (**Figure 2** shows the different ROIs). Using IBM SPSS (SPSS Inc., Chicago, IL, USA), we conducted repeated-measures ANOVA with within-group factors of group (children, adult), condition (Watch, Do, Together), hemisphere (left, right), and ROI (SA, IP, IA) and a between-group factor of group (child, adult) for average HbO$_2$ values. Greenhouse-Geisser corrections were applied when our data violated the sphericity assumption based on Mauchly's test of sphericity. For multiple *post hoc* comparisons, we have used the False Discovery Rate (FDR) method proposed by Singh and Dan (2006) for multichannel fNIRS data. We specifically used the Benjamin-Hochberg method wherein unadjusted p-values are rank-ordered from low to high. Statistical significance is declared if the unadjusted p-value is less than the p-value threshold. p-threshold was determined by multiplying 0.05 with the ratio of unadjusted p-value rank to the total number of comparisons (p-threshold for i$^{\text{th}}$ comparison = $0.05 \times i/n$; where n = total number of comparisons). Paired t-tests were used to examine group differences in behavioral data including temporal/spatial synchrony score, motor score, and additional movements.

RESULTS

Quality of IPS Behaviors

Paired t-tests showed that children had lower spatial synchrony scores (Mean ± SE = 2.67 ± 0.08) compared to adults (2.85 ± 0.06, $p = 0.03$). Similarly, children had lower temporal synchrony scores (2.74 ± 0.06) compared to adults (2.88 ± 0.05, $p = 0.04$) indicating more errors in the children vs. the adults. There were no significant group differences in terms of motor pattern errors (Children: 2.97 ± 0.01; Adults: 2.97 ± 0.01, $p > 0.1$) or additional movements (Children: 1.41 ± 0.43; Adults: 0.73 ± 0.23, $p > 0.1$; **Table 2**).

Cortical Activation During IPS

The group × condition × hemisphere × region four-way repeated ANOVA revealed a significant main effect of group ($F_{(1,119)} = 7.6$, $p = 0.007$), condition ($F_{(2,229.4)} = 145.2$, $p < 0.001$), hemisphere ($F_{(1,119)} = 30.7$, $p < 0.001$), and region ($F_{(1.8,220.2)} = 132.3$, $p < 0.001$) as well as two-way interactions between group × condition ($F_{(1.8,217.2)} = 17.6$, $p < 0.001$), group × hemisphere ($F_{(1,119)} = 12.4$, $p < 0.001$), condition × hemisphere ($F_{(1.7,203.5)} = 43.4$, $p < 0.001$), condition × region ($F_{(3.6,422.1)} = 18.2$, $p < 0.001$), as well as three-way interactions between group × condition × hemisphere ($F_{(1.4,170.5)} = 5.5$, $p = 0.01$) and condition × hemisphere × region ($F_{(3.3,396.2)} = 4.9$, $p = 0.04$). *Post hoc* analyses were focused on two three-way interactions of condition × hemisphere × region and group × condition × hemisphere (**Table 3** shows the Mean and SE of HbO$_2$ concentration values, **Table 4** shows significant p-values and direction of effects, and **Figure 7** shows channel specific activation data).

TABLE 2 | The quality of interpersonal synchrony (IPS) in the TD children and adults.

Video coding variables	Child (Mean ± SE)	Adult (Mean ± SE)
Spatial IPS	2.67 ± 0.08*	2.85 ± 0.06
Temporal IPS	2.74 ± 0.06*	2.88 ± 0.05
Motor score	2.97 ± 0.01	2.97 ± 0.01
Do condition	2.96 ± 0.02	2.97 ± 0.01
Together condition	2.99 ± 0.01	2.97 ± 0.01
Additional movements	1.41 ± 0.43	0.73 ± 0.23

**Indicates a significant difference between groups.*

Regional Differences in Both Groups

During the Watch condition, both groups had greater activation in the IA (STS) region (left and right ROIs) compared to SA (IFG) and IP (IPL) regions (ps ≤ 0.001). During the Do and Together conditions, both groups had greater SA (IFG) and IA (STS) activation compared to the IP (IPL) region for both left and right ROIs. Lastly, during the Together condition, both groups had greater activation in the right SA (IFG) than right IA (STS; ps < 0.01, **Figures 4**, **7**).

Task-Related Differences in Both Groups

Both children and adults showed greater activation over the left and right ROIs during the Do and Together conditions compared to the Watch condition (ps < 0.05, **Figures 5**, **7**). The differences between Do and Together conditions were only limited to the right hemisphere. Both groups showed greater activation in the right ROIs only (not left) during Together compared to the Do condition (ps < 0.05, **Figures 5**, **7**).

Hemispheric Differences in Both Groups

Adults had greater activation in the left hemisphere compared to the right hemisphere during both movement conditions of Do and Together (ps < 0.001, see **Figure 5**'s left vs. right comparisons and **Figure 7**). However, children had greater activation in the left hemisphere compared to the right hemisphere for the Do condition only ($p < 0.001$, see **Figure 5**'s left vs. right comparison and **Figure 7**) but not the Together condition ($p > 0.1$).

Group Differences Across Tasks

During the Watch condition, children had greater activation in the right hemisphere than the adults (ps < 0.001, **Figures 6**, **7**).

TABLE 3 | The mean and standard error (SE) of activation based on HbO$_2$ concentration values.

Group activation data	Watch		Do		Together	
	Mean	SE	Mean	SE	Mean	SE
TD child						
Left hemisphere						
SA/fronto-parietal	0.007	0.004	0.052	0.004	0.053	0.005
IA/temporal	0.020	0.006	0.055	0.006	0.052	0.007
IP/inferior parietal	−0.006	0.004	0.006	0.005	0.007	0.005
Right hemisphere						
SA/fronto-parietal	0.011	0.005	0.041	0.005	0.053	0.006
IA/temporal	0.032	0.007	0.030	0.006	0.040	0.007
IP/inferior parietal	−0.008	0.004	−0.003	0.004	0.002	0.005
TD adult						
Left hemisphere						
SA/fronto-parietal	0.001	0.004	0.079	0.010	0.076	0.009
IA/temporal	0.012	0.004	0.087	0.008	0.091	0.008
IP/inferior parietal	−0.009	0.003	0.051	0.006	0.045	0.006
Right hemisphere						
SA/fronto-parietal	0.003	0.004	0.038	0.007	0.061	0.007
IA/temporal	0.007	0.004	0.033	0.005	0.042	0.005
IP/inferior parietal	−0.013	0.003	−0.001	0.003	0.014	0.004

TABLE 4 | A listing of significant *p*-values and direction of the effect during *post hoc t*-tests.

Comparison	Significant *p*-values	Direction of effect
Group differences		
Watch, R hemisphere	<0.001	Child > Adult
Do, L hemisphere	<0.001	Adult > Child
Together, L hemisphere	<0.001	Adult > Child
Task-related differences		
Group × condition × hemisphere (regions were pooled)		
Adult, L hemisphere	<0.001	D & T > W
Adult, R hemisphere	<0.001	T > D > W
Child, L hemisphere	<0.001	D & T > W
Child, R hemisphere	<0.010	T > D > W
Condition × hemisphere × region (groups were pooled)		
Left SA, IA & IP	<0.001	D & T > W
Right SA, IA & IP	<0.05	T > D > W
Hemispheric differences		
Group × condition × hemisphere (regions were pooled)		
Adult, Do	<0.001	L > R
Adult, Together	<0.001	L > R
Child, Do	<0.001	L > R
Condition × hemisphere × region (groups were pooled)		
SA, IA, & IP ROIs for Do	<0.001	L > R
IA & IP ROIs for Together	<0.001	L > R
Regional differences		
Watch, L & R hemispheres	<0.010	IA > SA > IP
Do, L & R hemispheres	<0.001	SA & IA > IP
Together, L hemisphere	<0.001	SA & IA > IP
Together, R hemisphere	<0.010	SA & IA > IP

During the Do and Together conditions, adults had greater left hemispheric activation than children ($ps < 0.001$, **Figures 6**, **7**).

Correlation Between IPS Behaviors and Cortical Activation

For correlations between IPS behaviors and cortical activation, adults showed more significant correlations compared to children (number of correlations in children = 3 and adults = 12, **Table 5**). More specifically, the adult IPS spatial scores correlated with cortical activation in all three ROIs across all conditions. In addition, the adult IPS temporal scores correlated with SA and IP activation only during the Do and Together conditions. In slight contrast, the children's IPS spatial scores correlated with right SA and left IP activation in the Do and Together conditions and their IPS temporal scores did not correlate with cortical activation.

DISCUSSION

Previous fMRI studies of IPS have been limited to simple hand movements and unnatural environments. Using fNIRS, two studies have reported differences in cortical activation during IPS vs. solo action execution (Egetemeir et al., 2011; Bhat et al., 2017). However, no study has compared developmental differences in IPS performance as well as underlying cortical activation patterns between children and adults. The present study compared the cortical activation patterns between children and adults performing action observation, execution, and IPS during a naturalistic reach to the cleanup task. Consistent with our hypothesis, TD children had lower IPS than adults with children showing lower spatial and temporal synchrony scores compared to adults. However, the two groups did not differ in terms of motor pattern scores or other body movements. These findings suggest that while the accuracy of simple reaching motions was similar between adults and children, the ability to synchronize reaching motions with another individual was still developing between childhood and adulthood.

We found some similarities as well as differences between the cortical activation patterns of children and adults. First, both groups had greater cortical activation during the Do (execution) and Together (IPS) conditions compared to the Watch (observation) condition. More importantly, in the Together (IPS) condition, both groups had greater right hemispheric activation compared to the Do condition. In terms of regional similarities, in the Watch condition, both groups had greater activation in the IA (superior temporal or STS) region compared to the SA (inferior frontal or IFG) and IP (inferior parietal or IPL) regions for both hemispheres. During the Do and Together conditions, both groups had greater SA (IFG) and IA (STS) activation compared to the IP (IPL) region in both hemispheres. Lastly, during the Together condition, both groups had greater activation in the right SA (IFG) than the right IA (STS) region. However, we noted some differences in cortical activation patterns between children and adults. In terms of the within-group, hemispheric differences, adults had greater left hemispheric activation (than right) for both Do and Together conditions. However, in the TD children, this pattern was seen only for the Do condition with more bilateral activation in the Together condition. In terms of the between-group differences, during the Watch condition, TD children had greater right hemispheric activation compared to adults. Additionally, in the Do and Together conditions, the adults had greater left hemispheric activation than the TD children. Lastly, adult spatial synchrony scores correlated with cortical activation in all three ROIs and their temporal synchrony scores correlated with SA (IFG) and IP (IPL) activation only. In contrast, children's spatial scores correlated with the right IFG and left IPL activation but not their temporal scores.

IPS Improves Between Childhood and Adulthood

Children had lower spatial and temporal synchrony scores compared to adults suggesting lower IPS in children than adults. To our knowledge, only one study has compared developmental differences in IPS performance. During a joint drumming task, adult-adult dyads showed the highest levels of IPS and least within-individual, inter-limb variability followed by older child-child dyads and lastly the younger child-child dyads, who showed the lowest levels of IPS and greatest within-individual, inter-limb variability (Kleinspehn-Ammerlahn et al., 2011). It was posited that the greater variability in arm movements of younger children contributed to their action inconsistency and ability to synchronize with each other. Although we could not find other comparisons of IPS performance between children

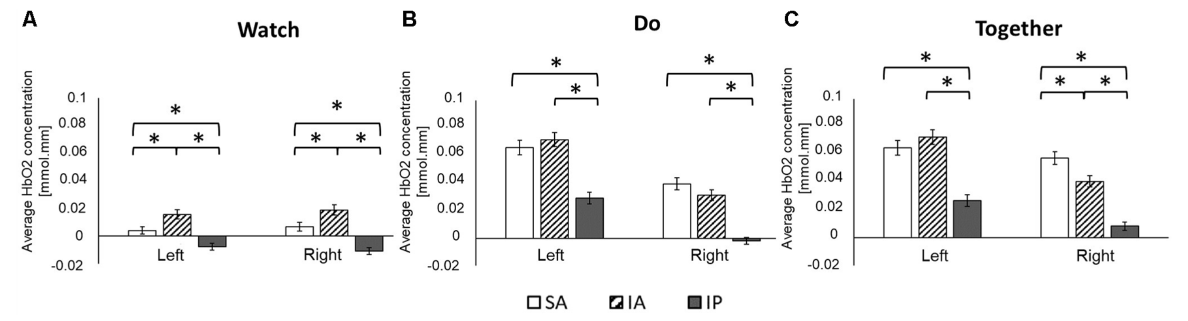


FIGURE 4 | Regional differences (**A**: Watch; **B**: Do; **C**: Together) in average HbO_2 concentration. *Indicates a significant difference between regions.

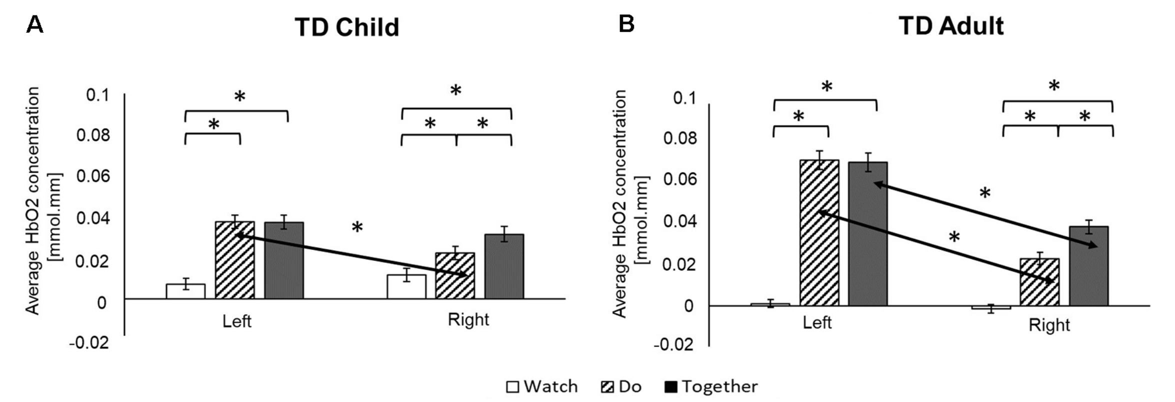


FIGURE 5 | Task-related and hemispheric differences for typically developing (TD) child **(A)** and adult **(B)** in average HbO_2 concentration. *Indicates a significant difference. Arrows highlight hemispheric differences.

and adults, one study comparing visual-motor synchronization of children and adults to various rhythmic visual stimuli found that 7–8-year-old children showed more variability and longer periods of asynchrony compared to adults (Kurgansky and Shupikova, 2011). Similarly, in our study, children made more errors in mirroring their choice of block or were more off in their timing of reaching or cleanup motions compared to the adults as they synchronized their actions to an adult tester. While our behavioral coding did not reveal any group differences in motor errors, we do not know if there was greater reaching variability in our child participants because we did not capture the reaching trajectories of both groups. Although behavioral coding did not reveal any obvious differences in the attentional patterns of both groups, possibly the differences in the visuomotor mapping of one's hand motions to that of the social partner could have contributed to the IPS differences between children and adults (Tahej et al., 2012).

STS Region Is Important for Observing Human Actions

During the Watch condition, both groups had greater STS activation than any other ROIs; however, children had greater right STS activation than the adults. The adult portion of this study was conducted before the child portion of the study. From coding of adult data, we noticed that mere instruction to watch during the Watch condition did not lead to careful observation of the tester's reaching motions. Hence, for the child group, in order to promote sustained attention, we asked the children to observe our motions carefully so that they could answer questions about how the task was completed at the end of the trial. Specifically,

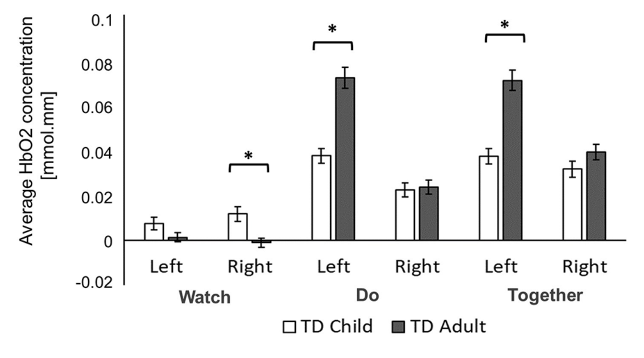


FIGURE 6 | Group differences in average HbO_2 concentration. *Indicates a significant difference between groups.

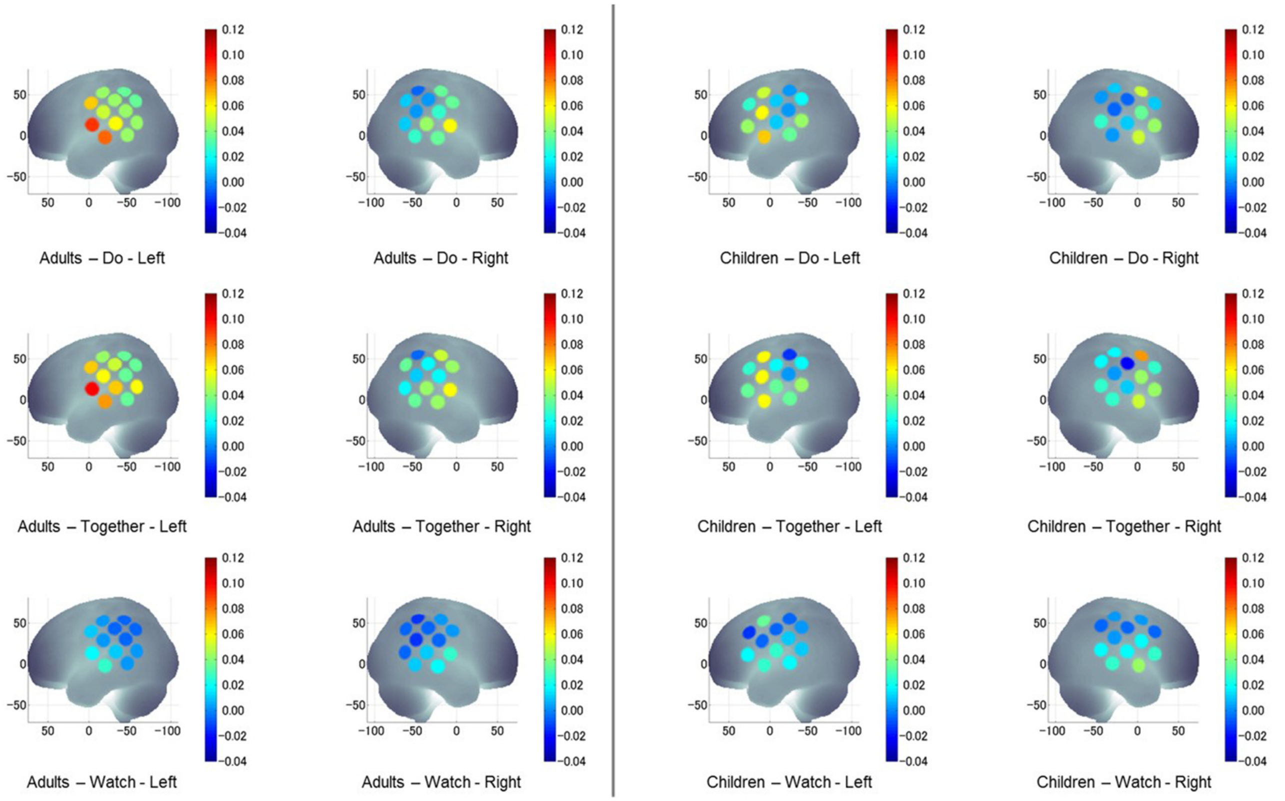


FIGURE 7 | A visual representation of task-related, hemispheric, and group differences in channel activation (HbO_2) during the stimulation period compared to its own baseline in both groups and all conditions.

TABLE 5 | Correlation between IPS behaviors and cortical activation.

		IPS spatial			**IPS temporal**		
		Watch	**Do**	**Together**	**Watch**	**Do**	**Together**
TD children							
SA	Left	ns	ns	ns	ns	ns	ns
	Right	ns	−0.36**	−0.40**	ns	ns	ns
IA	Left	ns	ns	ns	ns	ns	ns
	Right	ns	ns	ns	ns	ns	ns
IP	Left	0.23*	0.20*	0.38**	ns	ns	ns
	Right	ns	ns	ns	ns	ns	ns
TD adult							
SA	Left	ns	0.30**	0.33**	ns	0.42**	0.44**
	Right	0.24**	ns	ns	ns	−0.21*	−0.28**
IA	Left	ns	−0.21*	ns	ns	ns	ns
	Right	ns	0.27**	0.33**	ns	ns	ns
IP	Left	ns	0.27**	0.34**	ns	0.28**	0.33**
	Right	0.23*	ns	0.20*	ns	ns	ns

*The table presents r values using Spearman-rank correlations. **p < 0.01, *p < 0.05, ns: non-significant.*

we asked questions, for example, "Which block was cleaned up first or last, etc.?" This may have contributed to the greater social attention as well as greater STS activation observed in the children vs. the adults. Nevertheless, the results were similar between the two groups in that both groups had predominant STS activation during the Watch condition compared to activation in other ROIs (Bhat et al., 2017). Multiple fMRI studies have reported significant STS activation during action observation tasks (Montgomery et al., 2007; Molenberghs et al., 2010; Gatti et al., 2017). The STS region is important for processing and distinguishing social information such as biological motion, goal-directed actions of others, and mutual social gaze (Pelphrey et al., 2003a,b; Pelphrey and Morris, 2006). Pelphrey et al. (2003a) showed greater STS activation during observation of human or robotic motions compared to non-biological, object-related motions. Our common finding of greater fNIRS-based activation in superior temporal cortices during action observation in both groups (adults and children) is consistent with findings from past fMRI studies. Additionally, during a computerized ball toss game involving healthy adults, fNIRS-based activation was increased within the STS region when observing biological motion within a social inclusion context vs. a social exclusion context (Bolling et al., 2013). Therefore, in agreement with multiple fMRI and the few fNIRS studies (Bolling et al., 2013; Bhat et al., 2017) we also found greater STS activation during social observation of other's actions in both children and adults.

Role of IFG, STS, and IPL During Goal-Directed Actions and Their Importance for Visuomotor Correspondence During Imitation and IPS

During the movement conditions of the reach to cleanup task i.e., Do and Together conditions, both groups had greater activation in the IFG and STS compared to IPL regions. Moreover, adult synchrony scores correlated with cortical activation in all three ROIs whereas children's synchrony scores correlated with right IFG and left IPL activation only. Activation in the IFG was not very surprising because these regions are important for goal-directed movements (i.e., both Do and Together conditions required accurate reaching to objects; Cincotta and Ziemann, 2008). Similarly, during the self-selected motor task (i.e., the Do condition) we found temporal cortex activation (i.e., STS and middle temporal gyrus) in spite of no overt social interactions between the participant and tester. Testers were asked to avoid eye contact and overt social interactions with the participant during action execution. Additionally, we viewed the video data to remove any Do trials that involved social interactions; however, the mere presence of the tester may have contributed to some of the STS activations. Our findings somewhat fit with the current fMRI literature reporting significant STS activation during action imitation tasks compared to action execution and observation (Montgomery et al., 2007; Molenberghs et al., 2010). During object-based gesture tasks, bilateral STS activation was greater during action imitation compared to action execution and observation, which had similar activation levels (Montgomery et al., 2007). STS regions are said to provide a visual description of actions (Iacoboni, 2005). Molenberghs et al. (2010) suggested that STS is not merely registering the biological motions during imitation but also encoding the visuomotor correspondence between one's own actions and that of the partner. An fMRI study measuring cortical activation during observation of congruent vs. incongruent actions between two individuals revealed greater STS activation in the incongruent than congruent condition further corroborating the idea that STS may indeed be encoding visuomotor correspondence between individuals when moving together (Shibata et al., 2011).

The STS region could be interacting with IFG and other regions to receive efference copies of the motor plans to match the performed actions with the visual descriptions of imagined or observed actions (Iacoboni, 2005; Molenberghs et al., 2010). In our study, cortical activation during IPS was more similar to that of activation during action execution (not action observation). Additionally, synchrony errors in both groups correlated most with the Do and Together conditions (nine correlations per condition) and lastly the Watch condition (three Watch correlations). We believe our findings show that the challenges of imitation/IPS control stem from the complexity of motor components and not so much the observation component. It is often reported in the literature that simpler imitative tasks require less cortical activation compared to complex motor tasks and imitation performance is inextricably linked to its motor requirements such as body parts/joints involved as well as action complexity (Gatti et al., 2017).

The IPL region is also said to play an important role in planning the kinematics and goals of solo and complementary gestures/actions (Buxbaum et al., 2006; Sacheli et al., 2015). Specifically, the left parietal lobe contributes to visuospatial planning of actions with its lesions resulting in more errors during meaningless actions due to their more complex planning requirements (Tessari et al., 2007). Transcranial magnetic stimulation of dorsal parietal cortices interfered with online adjustments of reach-grasp actions suggesting its role in integrating end goals and motor planning (Tunik et al., 2008). Similarly, left parietal cortex activation was also reported by Sacheli et al. (2015) when performing joint grasping tasks in order to encode shared goals of complementary actions. Taken together, our findings fit with past literature confirming the role of STS, IFG, and IPL regions for visuomotor correspondence during both solo and synchronous reach-grasp actions.

Greater Left-Hemispheric Activation During Movement Tasks With Adults Showing More Left Lateralization

In general, during unilateral movement tasks of reach and cleanup, the two groups differed in that the adults had greater left lateralization than the children. To be clear, even when children used the right arm to complete the task there was perhaps some low-level "mirror" activation present in the homologous muscles of the left arm. This inability to suppress activity in the ipsilateral sensorimotor cortices has been reported in past studies comparing unilateral motor tasks between adults and children (Mayston et al., 1999; Huo et al., 2011). Mirror movements have been reported in children below 10 years of age but will diminish into adolescence and adulthood (Connolly and Stratton, 1968; Nass, 1985). Studies using transmagnetic stimulation (TMS) and magnetoencephalography (MEG) showed that during unilateral finger movements young children had activation in both contralateral and ipsilateral motor cortices due to lack of transcallosal inhibition resulting in muscle activity in the homologous muscles of the less active arm. However, this pattern of bilateral activation was not seen in adolescents and adults (Mayston et al., 1999; Huo et al., 2011). It is not surprising to see similar cortical activation patterns in our study since 65% of our child sample is below 11 years of age.

Greater Right Hemispheric Activation During IPS in Adults and Children

Both groups had increased right-hemispheric activation during the Together condition, compared to the Do condition, in spite of the right-handed nature of the reach-cleanup task. These results suggest that while there is left-lateralization during the Do condition (unilateral movements), IPS constraints led to more bilateral activation. These findings concur with a comprehensive meta-analytic review that showed activation of bilateral networks including frontal, premotor, parietal, and the temporo-occipital cortex during imitation (Caspers et al., 2010). Aziz-Zadeh et al. (2006) had participants observe, imitate, or execute unilateral finger movements with right or left hands to cues shown in the right or left visual field (hand moving or fixation cross). The imitation condition involved greater right inferior frontal and inferior parietal cortex activation in contrast to action execution, which mainly activated the contralateral primary visual and motor cortices. It was suggested that even during unilateral action imitation there is greater ipsilateral cortical activation compared to action execution, which is primarily contralateral in its control (Aziz-Zadeh et al., 2006). Similarly, Biermann-Ruben et al. (2008) found that imitation of biological hand movements led to greater right fronto-temporal activation compared to non-biological hand motions. Another group of fMRI studies has compared specular (mirrored—left-hand tester/video, right hand of subject) and anatomical imitation (both tester and subject use the identical hand for imitation, both use their right or left hands) and report greater bilateral or right hemispheric activation during specular imitation compared to anatomical imitation (Koski et al., 2003; Mengotti et al., 2015). In terms of fNIRS literature, few studies have reported greater bilateral activation during synchronous/cooperative actions with another partner (Egetemeir et al., 2011; Liu et al., 2015; Bhat et al., 2017). During a table-setting task, Egetemeir et al. (2011) reported greater activation in bilateral IPL regions during a joint action condition compared to the solo action or observation condition. Similarly, when two adults engaged in a cooperation game, the fNIRS patterns suggested that the partner who was actively following had greater right IFG activation compared to the partner who was a passive follower (Liu et al., 2015). In short, multiple studies have suggested that action imitation requires significant right/bilateral hemispheric activation beyond what is required during action execution. Consistent with the current literature, both children and adults in our study showed greater right/bilateral hemispheric activation during the IPS condition.

Mechanistic Framework for IPS

In this section, we highlight the common components across the different frameworks explaining the underlying processes associated with IPS behaviors (Semin and Cacioppo, 2009; Iacoboni, 2009; Pineda, 2009; Vesper et al., 2010, 2017). When engaging in IPS, each partner must understand the shared task goal as well as each of their individual roles in the task. While the overall goals are shared and similar; each partner's goals can be individual and distinct (Vesper et al., 2010). For example, when cleaning up blocks "together," the common goal was to move matching blocks in-synchrony; however, each participant still had to identify the appropriate block, pick, and place it in the container. In fact, it has been shown that partners will forego the quality of their own actions to complement and support the broader goal of moving with their partner (Schmitz et al., 2017). Some examples of how partners modify their actions for accomplishing the shared goal include changes in action speed or salience or workspace (Vesper et al., 2009, 2010, 2011; Schmitz et al., 2017). In terms of cortical regions, the IFG region is considered important for goal understanding during goal-directed actions such as reaching (Fontana et al., 2012). Additionally, the interactions between IFG and IPL regions are important for motor planning and sensorimotor representations for anticipatory control of actions (Koski et al., 2002). Second, during IPS, participants engage in visual monitoring of environmental cues (block color/shape, container location) as well as the partner's actions/social cues. These environmental and social cues will help in anticipating/predicting how to shape one's own actions in response to the partner's actions and environmental constraints (Semin and Cacioppo, 2009; Vesper et al., 2010). For example, the tester may begin to move their hand in the direction of a specific block and the child/adult monitoring the actions of the tester will pick up on these preparatory actions to accurately mimic the direction of tester's actions. Additionally, participants will engage in moment-to-moment visual/reactive adaptations to account for changes in environmental cues and any corrective adjustments made by the tester as they continue to move in-synchrony (Semin and Cacioppo, 2009). As mentioned earlier, the STS region plays an important role in establishing visuomotor correspondence and would be activated as partners utilize joint attention or shared gaze to accurately monitor and match their own actions to that of their partner's actions in a predictive or reactive manner (Molenberghs et al., 2010). It should be noted that the aforementioned regions do not work in isolation and are constantly interacting with each other and other cortical (e.g., primary, premotor and prefrontal cortices), and subcortical (e.g., cerebellum important for predictive control, etc.) structures (Gazzola and Keysers, 2009; Iacoboni, 2009; Caspers et al., 2010). Our current study findings fit with the above accounts in that both adult and child groups showed greater right STS and right IFG activation during the IPS conditions of the reach to cleanup task that also required greater spatial and temporal synchrony.

Study Limitations

This preliminary study has some limitations in the study design. In terms of study design, we were unable to compare the brain activation patterns between the tester and the child.

In the future, we plan to conduct a hyper scanning study to examine brain coherence between individuals engaging in IPS and imitation tasks. Additionally, there was some discrepancy in trials per condition completed by the adults (eight trials) vs. the children (six trials); however, we have calculated an average across trials for each condition and session. In terms of fNIRS data acquisition, we limited our analysis to 24 data channels over the lateral cortical surfaces and that did not capture prefrontal and motor cortex activity. Similarly, we were unable to implement the short-separation channels to account for skin-related hemodynamic responses as is implemented in other recent studies (Nguyen and Hong, 2016). Our subsequent studies have incorporated the full, 52-channel set up to collect lateral and prefrontal cortical activation. In terms of spatial registration, although we followed the international 10-20 system when placing probe sets, the variation of head size and probe placement could have added to the variability and inconsistency in the spatial registration of data channels. Finally, our study only reports the average hemodynamic responses; however, future studies should analyze detailed moment-to-moment changes in the hemodynamic response such as the initial dip in profile (see Hong and Naseer, 2016).

CONCLUSIONS

In conclusion, the quality of IPS in school-age children did not reach adult levels, although their accuracy of reaching or attentional patterns appear similar to those of the adults. fNIRS was able to detect the developmental changes in cortical activation. Our results suggested that there is a pattern of greater right hemispheric activation when engaging in IPS tasks suggesting that IPS is a more complex behavior (above and beyond action observation or action execution) as it requires greater bilateral cortical activation. Moreover, children had less lateralization compared to adults during unilateral reach-cleanup motions suggesting a lack of transcallosal inhibition in children compared to adults. Lastly, the quality of adult synchrony correlated with activation in various cortical regions whereas the quality of child synchrony only correlated with activation in few cortical regions (i.e., right IFG and left IPL) providing further evidence for developmental differences in synchrony performance and its underlying control. In summary, there is a clear developmental trajectory for IPS behaviors as well as associated cortical activation patterns between childhood and adulthood. In the future, we plan to use these normative patterns to further understand atypical IPS behaviors and atypical cortical activation in children and adults with Autism Spectrum Disorder, a population that is known to have difficulties with imitation and IPS including difficulties in social and motor performance.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Delaware's Human Subjects Review Board. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

W-CS contributed to this project through data collections, data analysis, and manuscript writing. MC, MH, and ST contributed to this project through recruitment, data collections, data analysis, and proofreading. KP was involved in the conception/planning phases of the project and assisted with proofreading. DT contributed to this project through the writing and implementation of his anchor-based spatial registration codes. He also assisted with the graphing of related data figures and writing of the spatial registration portion of the manuscript. AB provided oversight and help with recruitment, data collections, analyses, and manuscript writing.

FUNDING

This work was supported by the National Institutes of Health through a shared instrumentation grant awarded to the University of Delaware (Grant #: 1S10OD021534-01, PI: AB) and pilot award funding through an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health (U54-GM104941). AB also thanks the Dana Foundation for its support of this fNIRS project through a Clinical Neuroscience Award. In addition, this publication was made possible by the Delaware INBRE program, supported by a grant from the National Institute of General Medical Sciences—NIGMS (P20 GM103446) from the National Institutes of Health and the State of Delaware. The contents of this manuscript are solely the responsibility of the authors and does not necessarily represent the official views of NIH.

ACKNOWLEDGMENTS

We would like to thank all the children and families who participated in this study. We also thank Jeffrey Eilbott for his support in training AB in the use of Hitachi fNIRS technology during her sabbatical training period at the Yale Child Study Center. We also thank Dr. David Boas and his team at Boston University for training AB on the use of Homer-2 and in sharing the MATLAB codes/functions with our group. We would also like to thank undergraduate students, Jessica Gibbons and Kristen Arborio, from the University of Delaware for their help with data collection, behavioral coding, and data analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2020.00057/full#supplementary-material.

REFERENCES

Aziz-Zadeh, L., Koski, L., Zaidel, E., Mazziotta, J., and Iacoboni, M. (2006). Lateralization of the human mirror neuron system. *J. Neurosci.* 26, 2964–2970. doi: 10.1523/JNEUROSCI.2921-05.2006

Bhat, A. N., Hoffman, M. D., Trost, S. L., Culotta, M. L., Eilbott, J., Tsuzuki, D., et al. (2017). Cortical activation during action observation, action execution, and interpersonal synchrony in adults: a functional near-infrared spectroscopy (fNIRS) study. *Front. Hum. Neurosci.* 11:431. doi: 10.3389/fnhum.2017.00431

Biermann-Ruben, K., Kessler, K., Jonas, M., Siebner, H. R., Bäumer, T., Münchau, A., et al. (2008). Right hemisphere contributions to imitation tasks. *Eur. J. Neurosci.* 27, 1843–1855. doi: 10.1111/j.1460-9568.2008.06146.x

Bolling, D. Z., Pelphrey, K. A., and Kaiser, M. D. (2013). Social inclusion enhances biological motion processing: a functional near-infrared spectroscopy study. *Brain Topogr.* 26, 315–325. doi: 10.1007/s10548-012-0253-y

Brakke, K., Fragaszy, D. M., Simpson, K., Hoy, E., and Cummins-Sebree, S. (2007). The production of bimanual percussion in 12- to 24-month-old children. *Infant Behav. Dev.* 30, 2–15. doi: 10.1016/j.infbeh.2005.08.001

Brihmat, N., Tarri, M., Quidé, Y., Anglio, K., Pavard, B., Castel-Lacanal, E., et al. (2017). Action, observation or imitation of virtual hand movement affect differently regions of the mirror neuron system and the default mode network. *Brain Imaging Behav.* 12, 1363–1378. doi: 10.1007/s11682-017-9804-x

Buxbaum, L., Kyle, K., Tang, K., and Detrec, J. (2006). Neural substrates of knowledge of hand postures for object grasping and functional object use: evidence from fMRI. *Brain Res.* 1117, 175–185. doi: 10.1016/j.brainres.2006.08.010

Carpenter, M., Nagell, K., and Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monogr. Soc. Res. Child. Dev.* 63, 1–143. doi: 10.2307/1166214

Caspers, S., Zilles, K., Laird, A. R., and Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *NeuroImage* 50, 1148–1167. doi: 10.1016/j.neuroimage.2009.12.112

Cattaneo, L., and Rizzolatti, G. (2009). The mirror neuron system. *Arch. Neurol.* 66, 557–560. doi: 10.1001/archneurol.2009.41

Cincotta, M., and Ziemann, U. (2008). Neurophysiology of unimanual motor control and mirror movements. *Clin. Neurophysiol.* 119, 744–762. doi: 10.1016/j.clinph.2007.11.047

Clark, J. E., and Phillips, S. J. (1993). A longitudinal study of intralimb coordination in the first year of independent walking: a dynamical systems analysis. *Child Dev.* 64, 1143–1157. doi: 10.2307/1131331

Connolly, K., and Stratton, P. (1968). Developmental changes in associated movements. *Dev. Med. Child Neurol.* 10, 49–56. doi: 10.1111/j.1469-8749.1968.tb02837.x

Coren, S. (1992). *Handedness Questionnaire. The Left-Hander Syndrome: The Causes and Consequences of Left-Handedness.* New York, NY: Free Press.

De Renzi, E., Faglioni, P., Lodesani, M., and Vecchi, A. (1983). Performance of left brain-damaged patients on imitation of single movements and motor sequences. Frontal and parietal-injured patients compared. *Cortex* 19, 333–343. doi: 10.1016/s0010-9452(83)80004-5

Egetemeir, J., Stenneken, P., Koehler, S., Fallgatter, A. J., and Herrmann, M. J. (2011). Exploring the neural basis of real-life joint action: measuring brain activation during joint table setting with functional near-infrared spectroscopy. *Front. Hum. Neurosci.* 5:95. doi: 10.3389/fnhum.2011.00095

Fontana, A. P., Kilner, J. M., Rodrigues, E. C., Joffily, M., Nighoghossian, N., Vargas, C. D., et al. (2012). Role of the parietal cortex in predicting incoming actions. *NeuroImage* 59, 556–564. doi: 10.1016/j.neuroimage.2011.07.046

Gatti, R., Rocca, M. A., Fumagalli, S., Cattrysse, E., Kerckhofs, E., Falini, A., et al. (2017). The effect of action observation/execution on mirror neuron system recruitment: an fMRI study in healthy individuals. *Brain Imaging Behav.* 11, 565–576. doi: 10.1007/s11682-016-9536-3

Gazzola, V., and Keysers, C. (2009). The observation and execution of actions share motor and somatosensory voxels in all tested subjects: single-subject analyses of unsmoothed fMRI data. *Cereb. Cortex* 9, 1239–1255. doi: 10.1093/cercor/bhn181

Ge, S., Ding, M. Y., Zhang, Z., Lin, P., Gao, J. F., Wang, R. M., et al. (2017). Temporal-spatial features of intention understanding based on EEG-fNIRS bimodal measurement. *IEEE Access* 5, 14245–24258. doi: 10.1109/access.2017.2723428

Ge, S., Wang, P., Liu, H., Lin, P., Gao, J., Wang, R., et al. (2019). Neural activity and decoding of action observation using combined EEG and fNIRS measurement. *Front in Hum. Neurosci* 13:357. doi: 10.3389/fnhum.2019.00357

Hamzei, F., Vry, M. S., Saur, D., Glauche, V., Hoeren, M., Mader, I., et al. (2016). The dual-loop model and the human mirror neuron system: an exploratory combined fMRI and DTI study of the inferior frontal gyrus. *Cereb. Cortex* 26, 2215–2224. doi: 10.1093/cercor/bhv066

Holtzer, R., Izzetoglu, M., Chen, M., and Wang, C. (2019). Distinct fNIRS-derived HbO2 trajectories during the course and over repeated walking trials under single and dual-task conditions: implications for within session learning and prefrontal cortex efficiency in older adults. *J. Gerontol. A Biol. Sci. Med. Sci.* 74, 1076–1083. doi: 10.1093/gerona/gly181

Hong, K. S., Khan, M. J., and Hong, M. J. (2018). Feature extraction and classification methods for hybrid fNIRS-EEG brain-computer interfaces. *Front. Hum. Neurosci.* 12:246. doi: 10.3389/fnhum.2018.00246

Hong, K. S., and Naseer, N. (2016). Reduction of delay in detecting initial dips from functional near-infrared spectroscopy signals using vector-based phase analysis. *Int. J. Neural Syst.* 26:1650012. doi: 10.1142/s012906571650012x

Hu, X. S., Arredondo, M. M., Gomba, M., Confer, N., DaSilva, A. F., Johnson, T. D., et al. (2015). Comparison of motion correction techniques applied to functional near-infrared spectroscopy data from children. *J. Biomed. Opt.* 20:126003. doi: 10.1117/1.jbo.20.12.126003

Huo, X., Wang, Y., Kotecha, R., Kirtman, E. G., Fujiwara, H., Hemasilpin, N., et al. (2011). High gamma oscillations of sensorimotor cortex during unilateral movement in the developing brain: a MEG study. *Brain Topogr.* 23, 375–384. doi: 10.1007/s10548-010-0151-0

Huppert, T. J., Diamond, S. G., Franceschini, M. A., and Boas, D. A. (2009). HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Appl. Opt.* 48, 280–298. doi: 10.1364/ao.48.00d280

Iacoboni, M. (2005). Neural mechanisms of imitation. *Curr. Opin. Neurobiol.* 15, 632–637. doi: 10.1016/j.conb.2005.10.010

Iacoboni, M. (2009). Neurobiology of imitation. *Curr. Opin. Neurobiol.* 19, 661–665. doi: 10.1016/j.conb.2009.09.008

Isabella, R. A., and Belsky, J. (1991). Interactional synchrony and the origins of infant-mother attachment: a replication study. *Child Dev.* 62, 373–384. doi: 10.2307/1131010

Jones, S. S. (2007). Imitation in infancy: the development of mimicry. *Psychol. Sci.* 18, 593–599. doi: 10.1111/j.1467-9280.2007.01945.x

Khan, M. J., Ghafoor, U., and Hong, K. S. (2018). Early detection of hemodynamic responses using EEG: a hybrid EEG-fNIRS study. *Front. Hum. Neurosci.* 12:479. doi: 10.3389/fnhum.2018.00479

Kirschner, S., and Tomasello, M. (2009). Joint drumming: social context facilitates synchronization in preschool children. *J. Exp. Child Psychol.* 102, 299–314. doi: 10.1016/j.jecp.2008.07.005

Kleinspehn-Ammerlahn, A., Riediger, M., Schmiedek, F., von Oertzen, T., Li, S. C., and Lindenberger, U. (2011). Dyadic drumming across the lifespan reveals a zone of proximal development in children. *Dev. Psychol.* 47, 632–644. doi: 10.1037/a0021818

Klem, G. H., Lüders, H. O., Jasper, H. H., and Elger, C. (1999). The ten-twenty electrode system of the international federation. the international federation of clinical neurophysiology. *Electroencephalogr. Clin. Neurophysiol. Suppl.* 52, 3–6.

Koski, L., Iacoboni, M., Dubeau, M. C., Woods, R. P., and Mazziotta, J. C. (2003). Modulation of cortical activity during different imitative behaviors. *J. Neurophysiol.* 89, 460–471. doi: 10.1152/jn.00248.2002

Koski, L., Wohlschläger, A., Bekkering, H., Woods, R. P., Dubeau, M. C., Mazziotta, J. C., et al. (2002). Modulation of motor and premotor activity during imitation of target-directed actions. *Cereb. Cortex* 12, 847–855. doi: 10.1093/cercor/12.8.847

Kurgansky, A. V., and Shupikova, E. S. (2011). Visuomotor synchronization in adults and seven- to eight-year-old children. *Hum. Physiol.* 37, 526–536. doi: 10.1134/s0362119711050100

Liu, T., Saito, H., and Oi, M. (2015). Role of the right inferior frontal gyrus in turn-based cooperation and competition: a near-infrared spectroscopy study. *Brain Cogn.* 99, 17–23. doi: 10.1016/j.bandc.2015.07.001

Lloyd-Fox, S., Blasi, A., and Elwell, C. E. (2010). Illuminating the developing brain: the past, present and future of functional near infrared spectroscopy. *Neurosci. Biobehav. Rev.* 34, 269–284. doi: 10.1016/j.neubiorev.2009.07.008

Macrae, C. N., Duffy, O. K., Miles, L. K., and Lawrence, J. (2008). A case of hand waving: action synchrony and person perception. *Cognition* 109, 152–156. doi: 10.1016/j.cognition.2008.07.007

Marsh, K. L., Isenhower, R. W., Richardson, M. J., Helt, M., Verbalis, A. D., Schmidt, R. C., et al. (2013). Autism and social disconnection in interpersonal rocking. *Front. Integr. Neurosci.* 7:4. doi: 10.3389/fnint.2013.00004

Mayston, M. J., Harrison, L. M., and Stephens, J. A. (1999). A neurophysiological study of mirror movements in adults and children. *Ann. Neurol.* 45, 583–594. doi: 10.1002/1531-8249(199905)45:5<583::aid-ana6>3.0.co;2-w

Meltzoff, A. N. (2007). The 'like me' framework for recognizing and becoming an intentional agent. *Acta Psychol.* 124, 26–43. doi: 10.1016/j.actpsy.2006.09.005

Mengotti, P., Ripamonti, E., Pesavento, V., and Rumiati, R. I. (2015). Anatomical and spatial matching in imitation: evidence from left and right brain-damaged patients. *Neuropsychologia* 79, 256–271. doi: 10.1016/j.neuropsychologia.2015.06.038

Mogan, R., Fischer, R., and Bulbulia, J. A. (2017). To be in synchrony or not? A meta-analysis of synchrony's effects on behavior, perception, cognition and affect. *J. Exp. Soc. Psychol.* 72, 13–20. doi: 10.1016/j.jesp.2017.03.009

Molenberghs, P., Brander, C., Mattingley, J. B., and Cunnington, R. (2010). The role of the superior temporal sulcus and the mirror neuron system in imitation. *Hum. Brain Mapp.* 31, 1316–1326. doi: 10.1002/hbm.20938

Montgomery, K. J., Isenberg, N., and Haxby, J. V. (2007). Communicative hand gestures and object-directed hand movements activated the mirror neuron system. *Soc. Cogn. Affect. Neurosci.* 2, 114–122. doi: 10.1093/scan/nsm004

Mühlau, M., Hermsdörfer, J., Goldenberg, G., Wohlschläger, A. M., Castrop, F., Stahl, R., et al. (2005). Left inferior parietal dominance in gesture imitation: an fMRI study. *Neuropsychologia* 43, 1086–1098. doi: 10.1016/j.neuropsychologia.2004.10.004

Naseer, N., and Hong, K. S. (2015). fNIRS-based brain computer interfaces: a review. *Front. Hum. Neurosci.* 9:172. doi: 10.3389/fnhum.2015.00172

Nass, R. (1985). Mirror movement asymmetries in congenital hemiparesis: the inhibition hypothesis revisited. *Neurology* 35, 1059–1062. doi: 10.1212/wnl.35.7.1059

Nguyen, H. D., and Hong, K. S. (2016). Bundled-optode implementation for 3D imaging in functional near-infrared spectroscopy. *Biomed. Opt. Express* 7, 3491–3507. doi: 10.1364/boe.7.003491

Okamoto, M., Dan, H., Sakamoto, K., Takeo, K., Shimizu, K., Kohno, S., et al. (2004). Three-dimensional probabilistic anatomical cranio-cerebral correlation *via* the international 10–20 system oriented for transcranial functional brain mapping. *NeuroImage* 21, 99–111. doi: 10.1016/j.neuroimage.2003.08.026

Pelphrey, K. A., and Morris, J. P. (2006). Brain mechanisms for interpreting the actions of others from biological-motion cues. *Curr. Dir. Psychol. Sci.* 15, 136–140. doi: 10.1111/j.0963-7214.2006.00423.x

Pelphrey, K. A., Mitchell, T. V., McKeown, M. J., Goldstein, J., Allison, T., and McCarthy, G. (2003a). Brain activity evoked by the perception of human walking: controlling for meaningful coherent motion. *J. Neurosci.* 23, 6819–6825. doi: 10.1523/JNEUROSCI.23-17-06819.2003

Pelphrey, K. A., Singerman, J. D., Allison, T., and McCarthy, G. (2003b). Brain activation evoked by perception of gaze shifts: the influence of context. *Neuropsychologia* 41, 156–170. doi: 10.1016/s0028-3932(02)00146-x

Phillips-Silver, J., and Keller, P. E. (2012). Searching for roots of entrainment and joint action in early musical interactions. *Front. Hum. Neurosci.* 6:26. doi: 10.3389/fnhum.2012.00026

Pineda, J. (2009). *Mirror Neuron Systems.* Switzerland: Springer Publishing.

Pokorny, J. J., Hatt, N. V., Colombi, C., Vivanti, G., Rogers, S. J., and Rivera, S. M. (2015). The action observation system when observing hand actions in autism and typical development. *Autism Res.* 8, 284–296. doi: 10.1002/aur.1445

Rabinowitch, T. C., and Knafo-Noam, A. (2015). Synchronous rhythmic interaction enhances children's perceived similarity and closeness towards each other. *PLoS One* 10:e0120878. doi: 10.1371/journal.pone.0120878

Repp, B., and Su, Y. (2013). Sensorimotor synchronization: a review of recent research (2006-2012). *Psychon. Bull. Rev.* 20, 403–452. doi: 10.3758/s13423-012-0371-2

Reynolds, J. E., Billington, J., Kerrigan, S., Williams, J., Elliott, C., Winsor, A. M., et al. (2019). Mirror neuron system activation in children with developmental coordination disorder: a replication functional MRI study. *Res. Dev. Disabil.* 84, 16–27. doi: 10.1016/j.ridd.2017.11.012

Sacheli, L., Candidi, M., Era, V., and Aglioti, S. (2015). Causative role of left aIPS in coding shared goals during human-avatar complementary joint actions. *Nat. Commun.* 6:7544. doi: 10.1038/ncomms8544

Sato, H., Tanaka, N., Uchida, M., Hirabayashi, Y., Kanai, M., Ashida, T., et al. (2006). Wavelet analysis for detecting body-movement artifacts in optical topography signals. *NeuroImage* 33, 580–587. doi: 10.1016/j.neuroimage.2006.06.028

Schmitz, L., Vesper, C., Sebanz, N., and Knoblich, G. (2017). Co-representation of others' task constraints in joint action. *J. Exp. Psychol. Hum. Percept. Perform.* 43, 1480–1493. doi: 10.1037/xhp0000403

Semin, G., and Cacioppo, J. (2009). "From embodied representation to co-regulation," in *Mirror Neuron Systems*, ed. J. Pineda (Switzerland: Springer Publishing), 107–120.

Shattuck, D. W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K. L., et al. (2008). Construction of a 3D probabilistic atlas of human cortical structures. *NeuroImage* 39, 1064–1080. doi: 10.1016/j.neuroimage.2007.09.031

Shibata, H., Inui, T., and Ogawa, K. (2011). Understanding interpersonal action coordination: an fMRI study. *Exp. Brain Res.* 211, 569–579. doi: 10.1007/s00221-011-2648-5

Singh, A., and Dan, I. (2006). Exploring the false discovery rate in multichannel NIRS. *NeuroImage* 33, 542–549. doi: 10.1016/j.neuroimage.2006.06.047

Sofianidis, G., Hatzitaki, V., Grouios, G., Johannsen, L., and Wing, A. (2012). Somatosensory driven interpersonal synchrony during rhythmic sway. *Hum. Mov. Sci.* 31, 553–566. doi: 10.1016/j.humov.2011.07.007

Sutoko, S., Sato, H., Maki, A., Kiguchi, M., Hirabayashi, Y., Atsumori, H., et al. (2016). Tutorial on platform for optical topography analysis tools. *Neurophotonics* 3:010801. doi: 10.1117/1.nph.3.1.010801

Tachibana, A., Noah, J. A., Bronner, S., Ono, Y., and Onozuka, M. (2011). Parietal and temporal activity during a multimodal dance video game: an fNIRS study. *Neurosci. Lett.* 503, 125–130. doi: 10.1016/j.neulet.2011.08.023

Tahej, P. K., Ferrel-Chapus, C., Olivier, I., Ginhac, D., and Rolland, J. P. (2012). Multiple representations and mechanisms for visuomotor adaptation in young children. *Hum. Mov. Sci.* 31, 1425–1435. doi: 10.1016/j.humov.2012.02.016

Tessari, A., Canessa, N., Ukmar, M., and Rumiati, R. (2007). Neuropsychological evidence for a strategic control of multiple routes in imitation. *Brain* 130, 1111–1126. doi: 10.1093/brain/awm003

Tsuzuki, D., Cai, D. S., Dan, H., Kyutoku, Y., Fujita, A., Watanabe, E., et al. (2012). Stable and convenient spatial registration of stand-alone NIRS data through anchor-based probabilistic registration. *Neurosci. Res.* 72, 163–171. doi: 10.1016/j.neures.2011.10.008

Tunçgenç, B., and Cohen, E. (2016a). movement synchrony forges social bonds across group divides. *Front. Psychol.* 7:782. doi: 10.3389/fpsyg.2016.00782

Tunçgenç, B., and Cohen, E. (2016b). Interpersonal movement synchrony facilitates pro-social behavior in children's peer-play. *Dev. Sci.* 21:e12505. doi: 10.1111/desc.12505

Tunik, E., Ortigue, S., Adamovich, S., and Grafton, S. (2008). Differential recruitment of anterior intraparietal sulcus and superior parietal lobule during visually guided grasping revealed by electrical neuroimaging. *J. Neurosci.* 28, 13615–13620. doi: 10.1523/JNEUROSCI.3303-08.2008

Vesper, C., Abramova, E., Bütepage, J., Ciardo, F., Crossey, B., Effenberg, A., et al. (2017). Joint action: mental representations, shared information and general mechanisms for coordinating with others. *Front. Psychol.* 7:2039. doi: 10.3389/fpsyg.2016.02039

Vesper, C., Butterfill, S., Knoblich, G., and Sebanz, N. (2010). A minimal architecture for joint action. *Neural Netw.* 23, 998–1003. doi: 10.1016/j.neunet.2010.06.002

Vesper, C., Soutschek, A., and Schubö, A. (2009). Motion coordination affects movement parameters in a joint pick-and-place task. *Q. J. Exp. Psychol.* 62, 2418–2432. doi: 10.1080/17470210902919067

Vesper, C., van der Wel, R., Knoblich, G., and Sebanz, N. (2011). Making oneself predictable: reduced temporal variability facilitates joint action coordination. *Exp. Brain Res.* 211, 517–530. doi: 10.1007/s00221-011-2706-z

Volkmar, F. R., Sparrow, S. S., Goudreau, D., Cicchetti, D. V., Paul, R., and Cohen, D. J. (1987). Social deficits in autism: an operational approach using the vineland adaptive behavior scales. *J. Am. Acad. Child Adolesc. Psychiatry* 26, 156–161. doi: 10.1097/00004583-198703000-00005

Wiltermuth, S. S., and Heath, C. (2009). Synchrony and cooperation. *Psychol. Sci.* 20, 1–5. doi: 10.1111/j.1467-9280.2008.02253.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 24 March 2020
doi: 10.3389/fnhum.2020.00102

A Minimal Turing Test: Reciprocal Sensorimotor Contingencies for Interaction Detection

Pamela Barone[1,2]*, Manuel G. Bedia[3,4] and Antoni Gomila[1,2]

[1] Department of Psychology, University of the Balearic Islands, Palma, Spain, [2] Human Evolution and Cognition Group (EvoCog), University of the Balearic, IFISC, Associated Unit to CSIC, Palma, Spain, [3] Department of Computer Science and Systems Engineering, University of Zaragoza, Zaragoza, Spain, [4] Interactive Systems, Adaptivity, Autonomy and Cognition Lab, Aragón Institute of Engineering Research, University of Zaragoza, Zaragoza, Spain

In the classical Turing test, participants are challenged to tell whether they are interacting with another human being or with a machine. The way the interaction takes place is not direct, but a distant conversation through computer screen messages. Basic forms of interaction are face-to-face and embodied, context-dependent and based on the detection of reciprocal sensorimotor contingencies. Our idea is that interaction detection requires the integration of proprioceptive and interoceptive patterns with sensorimotor patterns, within quite short time lapses, so that they appear as mutually contingent, as reciprocal. In other words, the experience of interaction takes place when sensorimotor patterns are contingent upon one's own movements, and vice versa. I react to your movement, you react to mine. When I notice both components, I come to experience an interaction. Therefore, we designed a "minimal" Turing test to investigate how much information is required to detect these reciprocal sensorimotor contingencies. Using a new version of the perceptual crossing paradigm, we tested whether participants resorted to interaction detection to tell apart human from machine agents in repeated encounters with these agents. In two studies, we presented participants with movements of a human agent, either online or offline, and movements of a computerized oscillatory agent in three different blocks. In each block, either auditory or audiovisual feedback was provided along each trial. Analysis of participants' explicit responses and of the implicit information subsumed in the dynamics of their series will reveal evidence that participants use the reciprocal sensorimotor contingencies within short time windows. For a machine to pass this minimal Turing test, it should be able to generate this sort of reciprocal contingencies.

Keywords: Turing test, interaction, sensorimotor contingencies, reciprocity, perceptual crossing

Edited by:
Daniel Polani,
University of Hertfordshire,
United Kingdom

Reviewed by:
Maria Herrojo Ruiz,
Goldsmiths, University of London,
United Kingdom
Vasiliki Vouloutsi,
Institute for Bioengineering
of Catalonia (IBEC), Spain

Correspondence:
Pamela Barone
pamela.barone@uib.es

Specialty section:
This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

Received: *27 June 2019*
Accepted: *05 March 2020*
Published: *24 March 2020*

Citation:
Barone P, Bedia MG and Gomila A (2020) A Minimal Turing Test: Reciprocal Sensorimotor Contingencies for Interaction Detection.
Front. Hum. Neurosci. 14:102.
doi: 10.3389/fnhum.2020.00102

INTRODUCTION

Alan Turing proposed a famous test in order to study whether machines can exhibit intelligent behavior (Turing, 1950). In the so-called "Turing test," participants were challenged to tell whether they were interacting with another human being or with a machine. The interaction took place by means of exchanging computer screen messages between the human and the machine, both located in separate rooms. If participants cannot tell apart whether they are communicating with a machine or a human being, Turing reasoned, it must be because the machine exhibits intelligence.

However, the distant verbal conversation of the Turing test is a sophisticated form of interaction, quite different from more basic and typical social exchanges that normally take place among people. These basic forms of interaction are mainly face-to-face and embodied, context-dependent and based on the detection of reciprocal sensorimotor contingencies (Gomila, 2002). In our view, interaction detection in these cases requires the integration of proprioceptive and interoceptive patterns with sensorimotor patterns, within quite short time lapses, so that they are experienced as mutually contingent, as reciprocal. In other words, the experience of interaction takes place when sensorimotor patterns are contingent upon one's own movements, and vice versa. I react to your movement, you react to mine. When I notice both components, within appropriate time windows, I come to experience an interaction.

Interaction detection makes possible intersubjective experience. Intersubjectivity is the ability to engage in the mutual recognition of mental states without explicitly representing them (Trevarthen, 1979). In second-person interactions, we experience another person's mind (i.e., agency) in a direct, immediate, non-theoretical, and non-inferential way (Gomila, 2001, 2002; Pérez, 2013; Gomila and Pérez, 2017). This basic understanding is claimed to be the primary form of social cognition in human development (Reddy, 2008; Gomila, 2012); and contrasts with the detached, offline, and inferential way of intentional attributions required to interact in the classical Turing test.

Therefore, we designed a "minimal" Turing task to study how much information is required to detect these reciprocal sensorimotor contingencies and whether we resort to interaction detection –in this basic sense– to tell apart a human from a machine. Our minimal Turing test is inspired in a virtual and simple framework known as the "perceptual crossing" paradigm (Auvray et al., 2009). In the initial perceptual crossing scenario, two participants had to recognize each other in a common unidimensional, virtual space (a line 600 pixels long). They were located in two different rooms, in front of a computer and, while moving in this unidimensional virtual space, they encountered three agents: the avatar of another human, a shadow avatar of the human (also called a "mobile lure," which repeated the other participant's movements in another part of the virtual space) or a fixed object. Participants interacted by moving the computer mouse along the line and by receiving tactile stimulation when they crossed over one of the three agents. However, participants could not see the line, their cursor or the avatars that represented each type of agent at any time of the task. They were asked to detect –by clicking– when the tactile stimulation following a crossing had been produced by another human agent.

The main result of the study was that participants did not distinguish between the human and the shadow avatar: the probability of a player clicking the mouse when encountering the person or the mobile lure was not significantly different, although the players' partners were kept constant (Auvray et al., 2009). However, the correct discrimination between both agents emerged when the authors analyzed the interaction dynamics: As both players were mutually searching for one another, the encounters between the two participants were more frequent than encounters between the participant and the lure, providing an informational cue that might permit a more reliable interaction detection (Auvray and Rohde, 2012).

The perceptual crossing paradigm has been modified in different ways to study the dynamics of human social interactions. Lenay et al. (2011), for example, extended the results from Auvray's experiment into a richer, two-dimensional scenario, with similar results. On the other hand, Iizuka et al. (2009) expanded the original procedure to test whether pairs of individuals could figure out if an interaction was live or not through a different interface. By moving their fingers left or right in a tactile screen, participants received tactile vibrations when they touched another object in the virtual space. Participants were exposed either to a live interaction with another, always the same, subject or with a recording of a previous live interaction, and were asked to distinguish between them. Although they found it a hard task at the beginning and failed to recognize the two types of interactions, some pairs could develop a turn-taking and signaling strategy that helped them to succeed. However, just 4 out of 10 pairs achieved such a strategy only after tens of trials (Auvray and Rohde, 2012).

In fact, conscious recognition from the other partner only emerged in a version of the perceptual crossing in which participants received different sounds instead of tactile stimulation (Lenay and Stewart, 2012). In this experiment, in which each tone was associated with a type of object, participants could identify the source of the stimulation. The perceptual crossing paradigm has also been employed along the lines of Turing, in robotic and simulation experiments to model the dynamics of interaction detection (see Auvray and Rohde, 2012 for a review). On the other hand, a different strategy has been to develop a visual Turing test, where the interaction is based on joint attention (Pfeiffer et al., 2011).

Recently, an important modification of this framework was implemented by Bedia et al. (2014) who started considering the coupled dynamics of participants' activities during the test. They devised a program in which a human interacted either with another human agent or with the computer along 10 rounds. The computer could display a shadow avatar of one's own behavior or an agent with oscillatory movements. Similarly to the study of Auvray et al. (2009), each participant moved the mouse along an invisible line but he/she only heard a sound (instead of receiving tactile stimulation) when he/she crossed over another agent. At the end of each round, participants decided whether they had interacted with a human or with the machine, like in the Turing test.

Bedia et al. (2014) found, as in Auvray's original study, that the participants were not able of consciously distinguishing between the human avatar and its shadow, and that there was a difference in the pattern of interaction between both conditions. In this case, they found that the probability of having a new stimulation[1] within 0.5 s after a crossing did differentiate the type of agent involved. In other words, the implicit detection of social contingencies was made evident by a pattern of back and forth movements of both agents

---

[1] That is, the percentage of crossings that are followed by a second crossing.

around the same point, to generate overcrossings within the half-second time window. This pattern was uncovered by their analysis of the temporal structure of the interaction between two players. A fractal 1/*f* structure (called pink noise) at many timescales of the history of collective interactions emerged only within genuine social interactions (i.e., in the human vs. human case). Moreover, the largest values of the multifractal spectrum width also only appeared in human-human interactions. This distinctive pattern, that exclusively came out in interactions between two human beings, led researchers to argue that genuine social engagement might be better characterized by a structure of cross-scale interactions captured by analyzing fractal 1/*f* scaling and multifractal spectrum (Bedia et al., 2014).

To sum up, previous research with the perceptual crossing paradigm is somehow paradoxical: while it seems to provide a right approach to study interaction detection through social contingencies, it also comes short to prove that our judgments of interaction are based on the reciprocal contingencies detected. In most studies, participants failed to consciously distinguish the shadow agent from the other human participant in spite of evidence of their implicit discrimination in the dynamics of the interaction. The only study that found conscious detection of human interaction through this paradigm (Iizuka et al., 2009) required long series of interaction iteration between fixed pairs of participants, where less than half of participating pairs succeeded. This suggests that the strategies these pairs of participants developed were idiosyncratic, due to the fact that they kept playing with the same partner, rather than resorting to a basic process of reciprocal contingency detection.

In this paper, our main goal is to show that humans do detect interaction through social contingencies. Our second goal is to explore the reasons of the paradoxical results of previous research. In our view, it has to do with the fact that the standard perceptual crossing paradigm includes the three kinds of agents –human, mobile lure, fixed– in each trial, and unimodal feedback only.

As a matter of fact, the motivation to include the shadow agent in the design was meant to parallel the experiment originally devised by Murray and Trevarthen (1985) to study infants' ability to detect the interaction with an adult. In their pioneer study, they examined the quality of the social interaction between 2 and 3 month-old-infants and their mothers employing a double-video communication system. The baby and the mother were in different rooms and their behaviors were recorded. The infant faced a monitor, which displayed the behavior of his mother, and the mother saw, on her screen, her baby's behavior. Authors found that babies could distinguish, through their expressive behavior, when they were interacting with their mother from the condition in which they were shown exactly the same sequence but recorded from a previous trial. They reasoned that, in the former condition, babies were responding in real time thus perceiving the mothers' behavior contingently upon their own, while in the latter, babies lacked the power to influence the images. In the non-contingent scenario, intersubjectivity failed as infants could not engage in the reciprocity of facial expressions with their mothers and, as a result, they showed puzzlement, negative emotional reactions, and reduced eye contact[2]. In the perceptual crossing paradigm, the mobile lure was introduced to include this offline condition: an agent that behaves just like an interacting agent in its global trajectory, but non-contingently upon the behavior of the partner, as its movement depends on that of the human avatar. However, in the perceptual crossing paradigm all conditions are present in each trial, instead of distinguishing online and offline blocks (the only exception is Iizuka et al., 2009).

Therefore, in order to better match Murray and Trevarthen (1985) design what is needed is, first, a condition that closely resembles the offline condition implemented by those authors. This requires an agent with a trajectory identical to one exhibited by an interactive event correctly recognized as such by both participants, but recorded, so that this trajectory is not deployed contingently upon the moves of the participant. If participants decide whether they are interacting with a human on the grounds of the reciprocal sensorimotor contingencies detected, they will judge the offline condition as non-interactive. On the assumption that only human agents can interact in this paradigm, participants will judge that their partner is a human whenever they experience these reciprocal contingencies. And they will judge that their partner is a bot, whenever they do not. For this reason, in our "minimal Turing test" we included trials for each condition, instead of mixing them up.

However, things can be not that simple. For if the participant adopted a passive strategy, one of observing how the other avatar moves, the offline agent could elicit the illusion of interaction, as the pattern of feedback would be identical to a real interaction and participants could experience some reciprocal contingences when interacting with this recorded trajectory and judge that it is a human. Participants should adopt an active strategy in order to perceive the others' movements as contingent to their own and vice versa. Hence, we tried to provide instructions to participants that fostered this active stance.

On the other hand, Murray and Trevarthen's (1985) design involved audiovisual contingencies. Previous research with the perceptual crossing paradigm already showed that auditory feedback was more discriminant than tactile stimulation to detect the relevant contingencies. But it may also be the case that the difficulty in consciously distinguishing real interaction from interaction with the mobile lure was because auditory information is not robust enough. Therefore, in our study we also wanted to explore the question of whether the minimal sensory contingencies need to involve more than one modality; in particular, whether audiovisual information is required for robust interaction detection, as in Murray and Trevarthen's (1985) study. In addition, we also wanted to explore whether participants improve their performance in the auditory feedback block after undergoing the audiovisual block. Thus, we first presented only auditory information about the interaction, then audiovisual stimulation and finally just auditory feedback again, to check whether participants' performance improved along the task.

---

[2] Other studies have also acknowledged that babies seem to use social contingency to recognize an interactive partner (Bigelow et al., 1996; Nadel et al., 1999; Bigelow and Decoste, 2003; Stormark and Braarud, 2004; Field et al., 2005; Okanda and Itakura, 2005)

In summary, in our minimal Turing test, we modified Bedia et al. (2014) version of the perceptual crossing paradigm to test participants separately in each condition, and through these different feedback blocks. In the first study, participants were exposed either to movements of a human agent, which could be online or offline, or to the movements of a computerized oscillatory agent in three different blocks. In each block, either auditory or audiovisual feedback was provided along each trial, to inform participants when they crossed over another agent. Following previous studies, we first analyzed participants' recognition about the nature of the other agent by paying attention to their explicit answers. Secondly, we investigated the implicit information subsumed in the interaction dynamics of each participant's series, like the correlations between the series of two interactors, the time between two crossings, the window of crossings (density of crossings), and the fractal indices, in order to find out whether they tried to solve the task through the crossing patterns they detected. Finally, we asked the participants about their experience with the task. In a second, follow-up study, we simplified the task by eliminating the oscillatory bot, just to focus on the interaction dynamics. We hypothesized that participants are able to detect the social contingencies and to use them to respond to the Turing test question, but wanted to explore which informational conditions turn out to be discriminant enough. Were the information available to the participant insufficient, they would adopt an observational attitude to answer the test question.

STUDY 1

Materials and Methods

Participants

A total of 70 participants (55 females) took part in this experiment. Their ages ranged from 20 to 48 years ($M = 23.15$ and $SD = 4.62$). They were recruited from a Psychology class at the University of the Balearic Islands. Participants received credit points for participating in the study. In the lab, they were arranged in groups of six people (11 groups) or in groups of four people (only one group).

Experimental Procedure

Each participant entered into a cubicle, wore headphones and sat in front of a computer. As we have six cubicles in the lab, the maximum number of participants that could do the test at the same time was six. Inside the cubicle, they could not see nor hear the other participants. They were instructed on the goal of the study: they had to move their computer mouse in a shared perceptual environment and they were going to hear a sound whenever they crossed over another agent. They had to detect, in each round of the experiment, whether they have interacted with a human or with a computerized agent (for an example of the procedure, see **Supplementary Material**).

The shared perceptual environment was a virtual, one-dimensional space 800 pixels long (a line) with both ends connected, forming a torus to avoid the singularities induced by the edges, as in Auvray's original study. Although all movements were available with the computer mouse, only movements from left to right (and vice versa) were considered by the software. When the cursor of a participant crossed the cursor of another agent –either a human or a computerized agent– a collision was perceived because they received an audible stimulus lasting 500 ms. Such audible stimuli were the only environmental perception during some block of trials while the computer screen was black along the whole round. In other block of trials, participants could also see the line and the avatars of each agent.

Specifically, participants started with a training phase in which they could see the line on the screen together with their own avatar (that represented their movements) and the partner's avatar on it. They performed 4 training rounds and each lasted 15 s. After that, participants received three blocks of experimental rounds. Each block consisted of 9 rounds and each round lasted 30 s. In the first block, participants could not see the line nor the avatars on the screen and just received auditory stimulation when they crossed over another agent. During the second block, participants received audiovisual information in each of the 9 rounds –as in the training phase. Finally, participants received a third block with only auditory feedback (identical to block one).

We decided to provide audiovisual stimulation in the training phase so participants could familiarize with the setup more easily, as they had two sources of environmental stimulation. Block 1 and block 2 were thought to measure how different amount of information would affect such discrimination. Block 3 finally aimed to check whether participants could improve their performance after the second, audiovisual, block.

In each round of the task, participants could encounter one of three possible agents.

1. A "human online agent," that is, one of the participants from another cubicle. As the task consisted of many trials, the participant was randomly assigned to any other human participant from the group of six to just focus on reciprocal contingency detection as a basic process, and avoid the development of idiosyncratic strategies.
2. A "human offline agent" that consisted in a recording of a previous human-human interaction in which both participants recognized each other as humans. The human offline agent was randomly chosen from an array of series in which two human players previously interacted and correctly recognized each other. Those series were collected from a pilot study and here presented offline.
3. An "oscillatory agent" that exhibited "a sinusoidal behavior (describing a sinusoidal trajectory of 0.5 Hz and 200 pixels of amplitude), predictable and deterministic" (as in Bedia et al., 2014, 4).

As we stated in the instruction, participants were asked to say whether they have interacted with a human or with a computerized agent. In fact, the computer could randomly present either the human offline agent or the oscillatory bot. Therefore, they were unaware of the nature of the computerized agents. The underlying assumptions were that if they experienced an interaction they would respond "person," that the offline agent would not generate the experience of interaction, and that if the experience of interaction failed to emerge, they would decide from an observational stance.

In the training phase, we only used two types of agents: the human online and another non-reactive bot, which increased its position with a velocity of 30 pixels per second during the whole round. We intentionally employed a different computerized agent that the ones that will appear on the test because we only aimed at introducing the movements of the "machine bots" without showing them the same behavior that were to appear in the task, to avoid any possible habituation or anticipation effect.

In each round of the experiment, participants moved their avatar (by moving their computer mouse) along the unidimensional space, and they heard a sound when they crossed over the other agent or when the other agent crossed over them. At the end of each round, participants had to answer the following question "Who have you interacted with in the last round?" by clicking on one of the answers: "person" or "machine." Using only auditory information (blocks 1 and 3) or using audiovisual information (training phase and block 2), they had to decide, at the end of each round, whether they interacted with a person or a computerized agent. After they answered, they received feedback about their choice (correct/incorrect). After completing all the procedure, participants were asked about their experience.

Coding

On the one hand, we coded participants' explicit responses at the end of each round as well as the total number of crossings.

Participants' responses

We coded participants' correct answers at the end of each round, that is, whether they correctly guessed the nature of the other agent. A correct answer was considered when they replied "person" after interacting with the human online agent, and "machine" in the rounds when they interacted with both computerized agents (human offline and oscillatory agent).

Number of crossings

We calculated the total number of crossings that each participant executed per round (i.e., the "active crossings," referring only to those crossings produced by the participant).

On the other hand, we coded different implicit measures in the dynamics of the interaction.

Fractal indices

A fractal index was obtained as in Bedia et al. (2014). In a nutshell, we first took the time series of the distance between the two agents in a round. We then computed the agents' relative velocity (i.e., the first derivative of the distance between the participant and the other agent) to extract whether they are approaching or distancing themselves at each moment of time. Then we used a detrended fluctuation analysis (DFA) algorithm (Peng et al., 2000) to compute the statistical self-affinity in the data series of distance variations (more detail description of the method can be found in the **Supplementary Material**; also see Bedia et al., 2014). The result is a Beta index (β) that characterizes distinct processes. Values of β close to 0 feature uncorrelated processes (correspond to white noise), values close to 2 exemplify rigid and deterministic processes (brown noise) and values of β close to 1 characterize flexible and adaptable processes (pink noise), that is, processes that are not totally organized but neither totally disorganized.

These background noises refer to intrinsic sources of variability, the intrinsic dynamics of mind and body, and the internal workings of a living being (Van Orden et al., 2003). The interest of this measure is that pink noise has been encountered in biological, physical and cognitive systems and is proposed as a signature of dynamic complexity (Gilden, 2001). These systems are sustained by interaction dominant dynamics; which consist of multiplicative interactions that imply coordination between the different timescales in the system (Van Orden et al., 2005). In the present study, pink noise would only emerge when the participant interacted with the human online, since the mutual detection of sensorimotor contingences would give rise to flexible and adaptable behavior from the partner. Brown noise would emerge when the participant interacted with the oscillatory bot, due to the deterministic movements of this agent; while white noise would better characterize the offline agent since no mutual detection of sensorimotor contingences can emerge along trials.

β was calculated for each series. As a result, we obtained a distribution of "how many β" were for each value within the interval (0, 2) for all the agents.

Time between two crossings and window of crossings

We estimated the time it takes to produce another crossing after one took place. This means that, for each crossing in the series, we determine the time until the following one occurs. For example, if there have been crossings at 1000, 3000, 3600, 3800, 3900, 7000, 7300, and 9000 ms of one round, then, the time between crossings would be 2000, 600, 200, 100, 3100, 300, and 1700 ms.

Relatedly, the window of crossing refers to the number of crossings counting after one crossing is produced up to a certain time after this crossing. In a sense, it is the density of crossings (crossings/time; see **Supplementary Material** for an example of this calculation). It was assumed that the detection of reciprocal contingencies would require rapid back and forth movements around the crossing point by the two partners.

Similarity between two series

We calculated the similarity between two series with the crossed-correlation function. This function compares two temporal series and returns a value. The bigger the value, the greater the similarity between both trajectories.

We applied the crossed-correlation function for different time gaps and window spans. A time gap implies comparing two trajectories applying a temporal delay to one of them, to check whether the similarity between both occurred within a specific delay (like an echo). Window spans were chosen based on the crossings. This means that, for each crossing in each round, we took the positions of the two players from their series and calculate their auto-correlation. The correlation of the round is the average of the correlations among all their crossings. Again, the reason of this measure was the expectation that in interactive trials the trajectories of the agents would be similar and the correlation would be higher.

Statistical Approach

The design of this experiment involved repeated measures per participant and, in order to account for this characteristic, linear mixed effect models was computed. In order to determine our dependent variable (successful reply), we introduced many predictors: type of agent (human online, human offline, or oscillatory agent); type of block (block 1, block 2, and block 3), participant's age, gender, β, crossings, and correlation (similarity) between two series.

To avoid a 'multiple comparisons' problem, we used bootstrapping tools. In short, bootstrapping methods take data to create new simulated models, providing more robust analyses as well. It has been proved as a generalization of classical multiple comparisons procedures (Berkovits et al., 2000) and its use in some multiple contexts works better than classical methods when it is required a large sample theory and to make the arguments in small samples (Gelman et al., 2012). As noted by Westfall (2011) the bootstrap provides a "simple, elegant generalization of classical multiple comparisons procedures" (p. 1188).

Results

Participants' Responses and Number of Crossings

Participants greatly succeeded in recognizing the oscillatory bot as a machine in all blocks (binomial two-sided tests, $ps < 0.001$ in the three blocks). Secondly, they recognized the human online as a person in the first two blocks (binomial two-sided tests, block 1: $p = 0.02$, block 2: $p < 0.001$) but not in block 3 ($p = 0.08$). When participants interacted with the human offline agent, they consistently replied that it was a person (binomial two-sided tests, all $ps < 0.005$; see **Table 1**).

The highest values of recognition appeared in block 2 when participants could also see the avatars on the screen (70 and 95% of successes for online and oscillatory agents, respectively). The comparison of performance in block 3 with block 1 (the blocks providing only auditory information about the crossings) revealed higher recognition of the oscillatory bot (from 66 to 70%) and of the offline agent (from 32 to 40%), while the number of correct answers in the condition of the human online agent barely decreased (from 59 to 56%), but any difference between block 1 and block 3 was significant.

Table 1 also shows the mean number of crossings per block and type of agent. For all type of agents, there were more crossings in the first block than in the other blocks, suggesting an exploratory strategy. In the first block, there were more crossings when the partner was the human online agent than if it was an offline agent or the oscillatory agent. In block 2, crossings decreased in general, but much more when the partner was the oscillatory agent. In block 3, crossings when the partner was human online decreased, and increased for the other two kind of agents. However, crossings did not significantly differ among the types of agent in each block either.

Figure 1 shows the probability of success and crossings.

Fractal Indices

Figure 2 shows a graph with the distribution of the values of β per agent and block. Values of β in human-human online rounds were lower in our study, along the three blocks, than the value reported in Bedia et al. (2014; β = 0.86). On the other hand, the β-values in human-human offline series were similar to the β that emerged in the human-shadow condition of Bedia et al. (2014; β = 0.29). In fact, β-values from both human-human online and human-human offline interactions were very similar in our study, and closer to 0 which characterizes uncorrelated processes (white noise) (see **Figure 2**, at the bottom). For a table with all the values of β per agent and block, see **Supplementary Material**.

TABLE 1 | Probability of success and total number of crossings in Study 1.

	BLOCK 1	BLOCK 2	BLOCK 3
Correct answers			
Online	0.59	0.70	0.56
Offline	0.32	0.28	0.40
Oscillatory bot	0.66	0.95	0.70
Number of crossings			
Online	37.08 (38.7)	27.89 (27.85)	25.71 (24.74)
Offline	32.64 (29.72)	28.38 (26.70)	27.61 (26.22)
Oscillatory bot	32.61 (24.48)	20.80 (27.71)	29.92 (28.19)

Probability of correct responses about the nature of the other agent and mean (and standard deviation) of crossings in each block and type of agent in Study 1.

Time Between Two Crossings and Window of Crossings

Figure 3 illustrates the distribution of crossings according to the time interval. Graphs on the top show the mean number of crossings, produced after one crossing of reference, in a definite amount of time (considering the previous crossing as a reference). As expected, most crossings took place in between 0.1 and 1 s after the previous crossing (the center of the Gaussian distribution is between 10^2 and 10^3 ms).

Graphs at the bottom measure how the crossings were distributed along the round. They represent the accumulated sum of the graphs on top and show that most of the crossings from a round -after one crossing of reference- were distributed in a window span of 1 s (after the initial crossing), generating "gusts of crossings." In the time interval from 1 to 10 s, the number of accumulated crossings was lower (the sigmoid curve reached a plateau).

The online agent, in Block 1, accumulated more crossings – after one of reference– in the time window of 100 ms than the rest of the agents. The accumulated crossings, after a previous one, sharply decreased in block 3 for the online agent, suggesting a change in how participants moved the cursor.

Figure 4 reports the distribution of crossings as we increase the window span, between 500 and 1500 ms. In general, the number of crossings diminished for the online agent from block 1 to block 3 (in line with the findings from the previous graph). In block 1, when a participant crossed over the online agent, the following crossings increased up to seven within the 1500 ms window span while, in block 3, they added up to four more crossings. The opposite pattern followed the distribution of crossings for the offline and oscillatory agents, since the crossings after the first one occurred slightly increased from block 1 to block 3 (from five to six crossings in the 1500 ms window span).

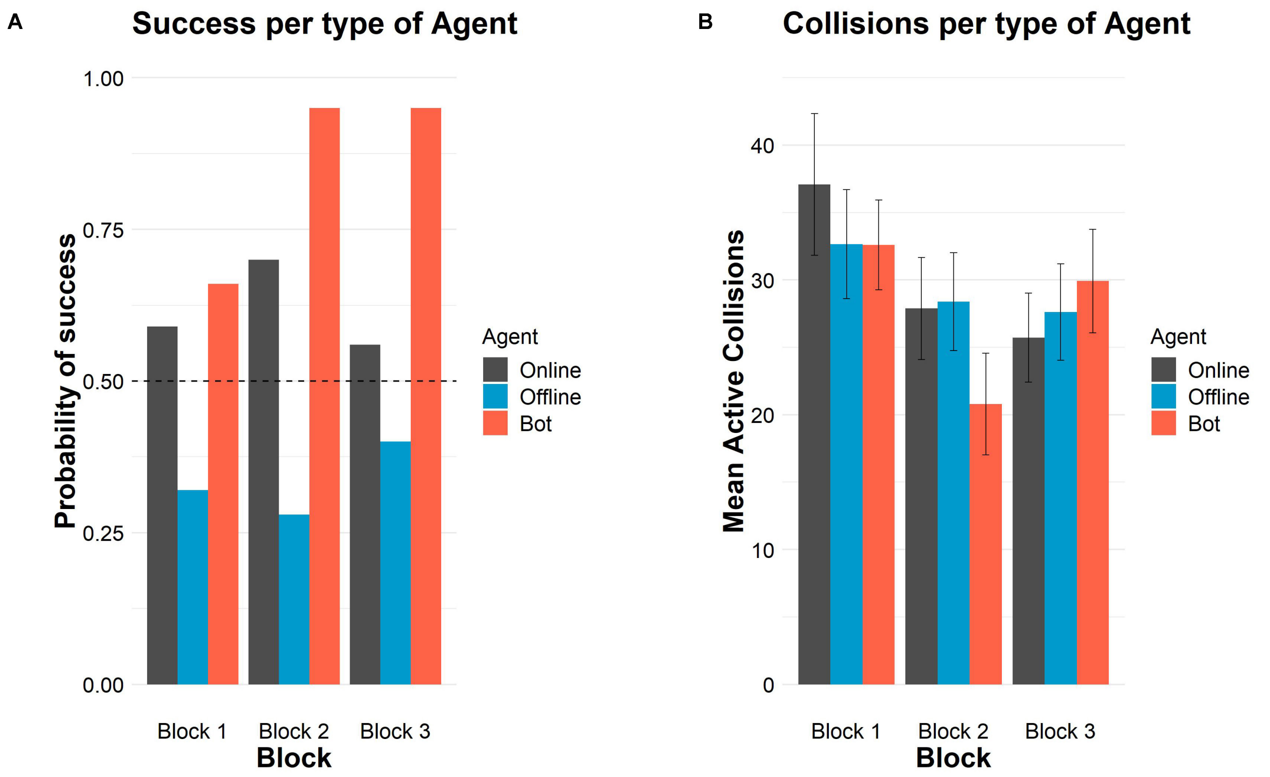


FIGURE 1 | Successes and crossings per type of agent and block in Study 1. **(A)** Probability of success in each block per type of agent. The horizontal dashed line represents chance level (50%). **(B)** Mean number of crossings in each block per type of agent. Error bars depict 95% confidence intervals.

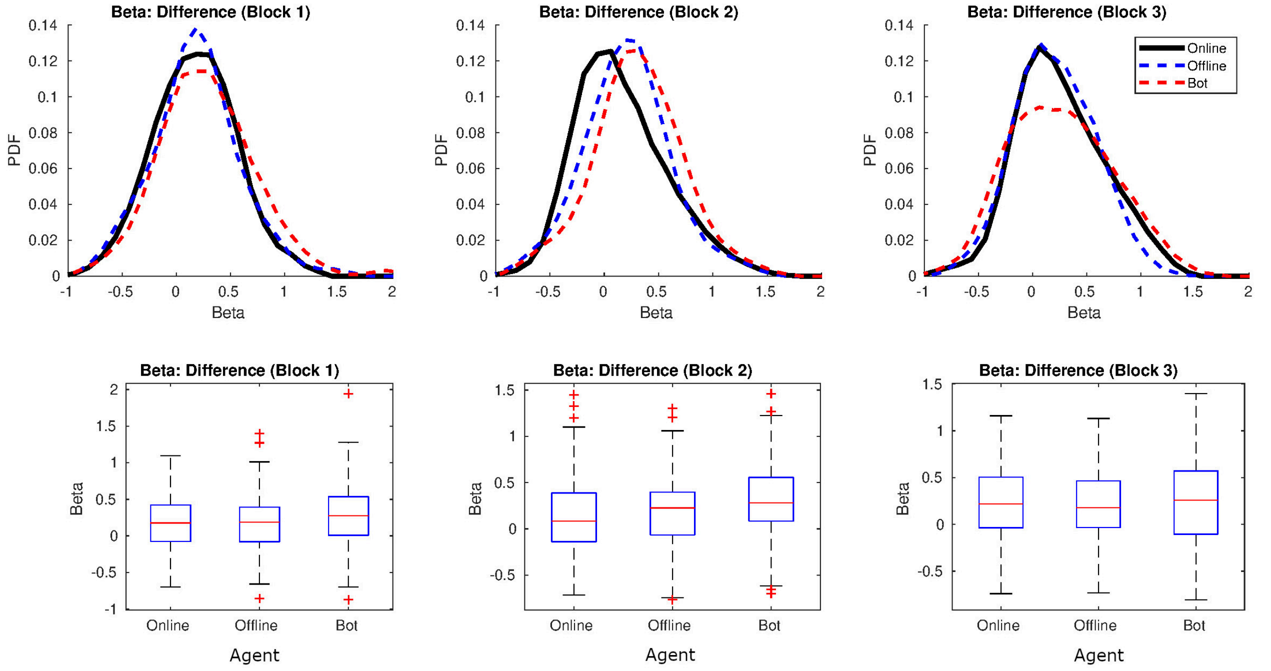


FIGURE 2 | Distribution of β in Study 1. On the top: distributions for the values of β (in black online agent, in blue offline agent, in red the oscillatory bot). On the bottom, representation of the means and variances as boxplots.

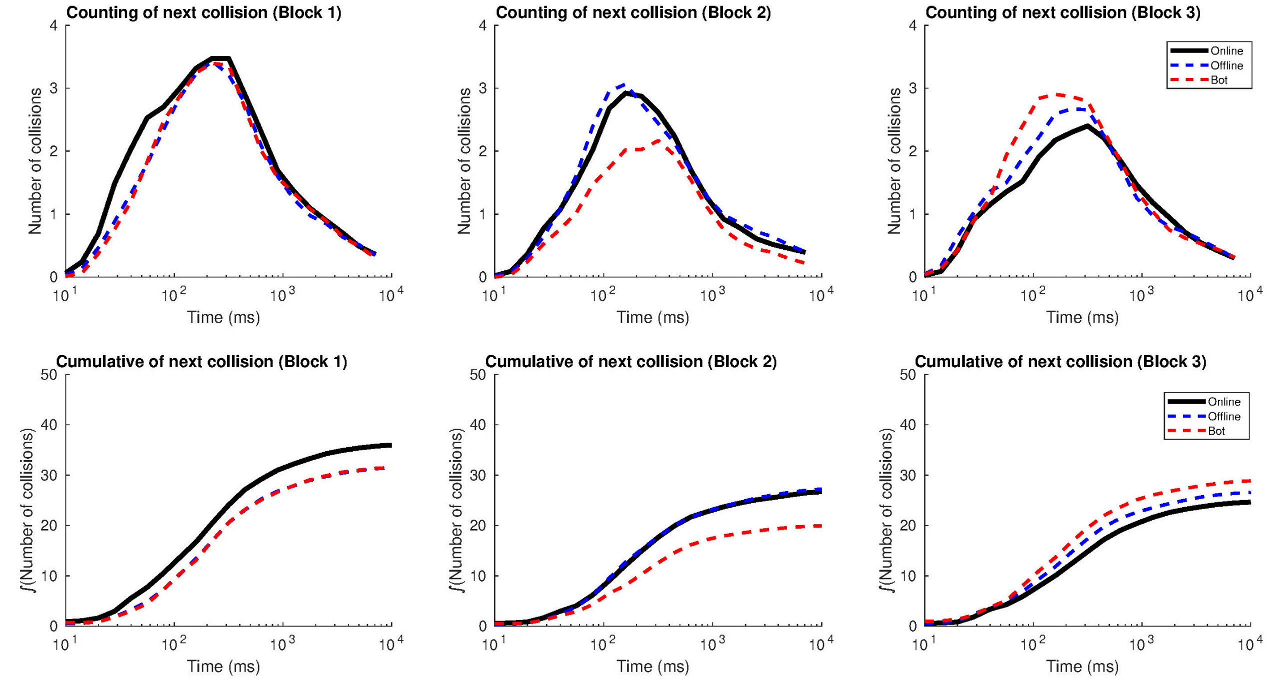


FIGURE 3 | Distribution of the number of crossings according to the window span in Study 1. The distribution has a Gaussian shape. The *x*-axis (time) is shown on a logarithmic scale in order to see the relation between short times (milliseconds) and longer ones (dozens of seconds, which is the length of a round). On the top: number of crossings after one crossing of reference in relation to the time they are produced, for each block and type of agent. On the bottom: number of accumulated crossings as a function of time.

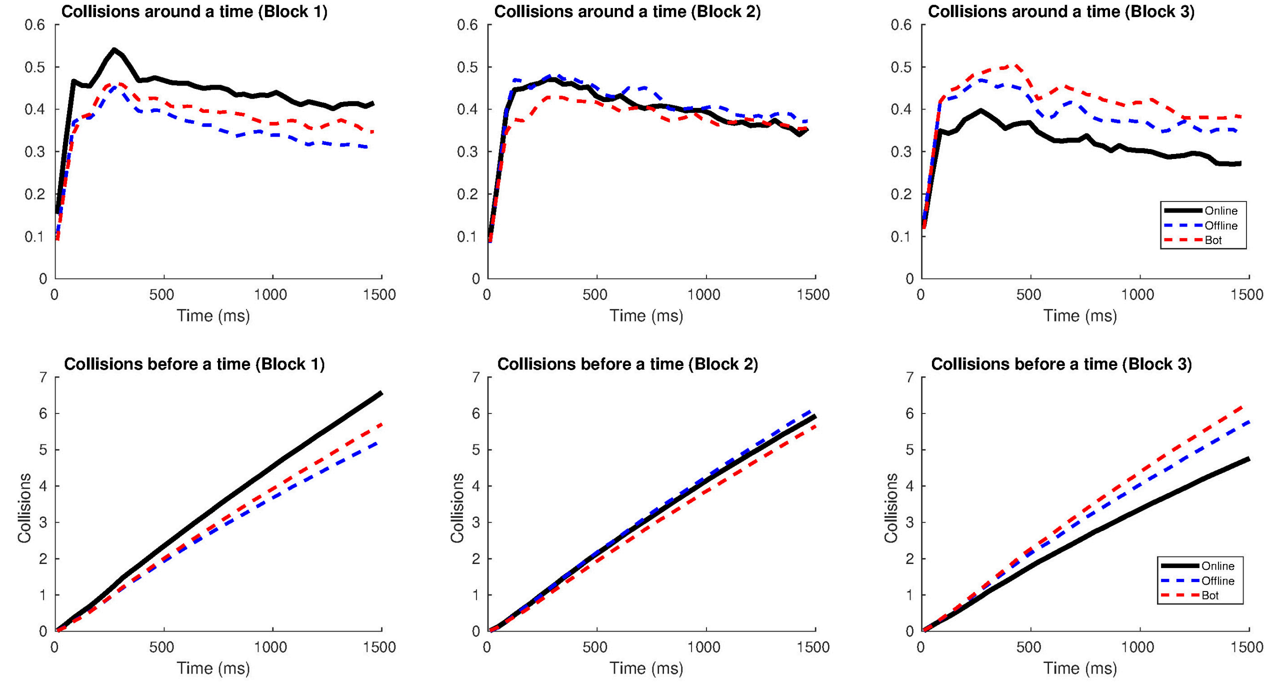


FIGURE 4 | Distribution of the number of crossings as we increase the window span in Study 1. The graphs show how many crossings are produced in the window span from one crossing until a specific amount of time.

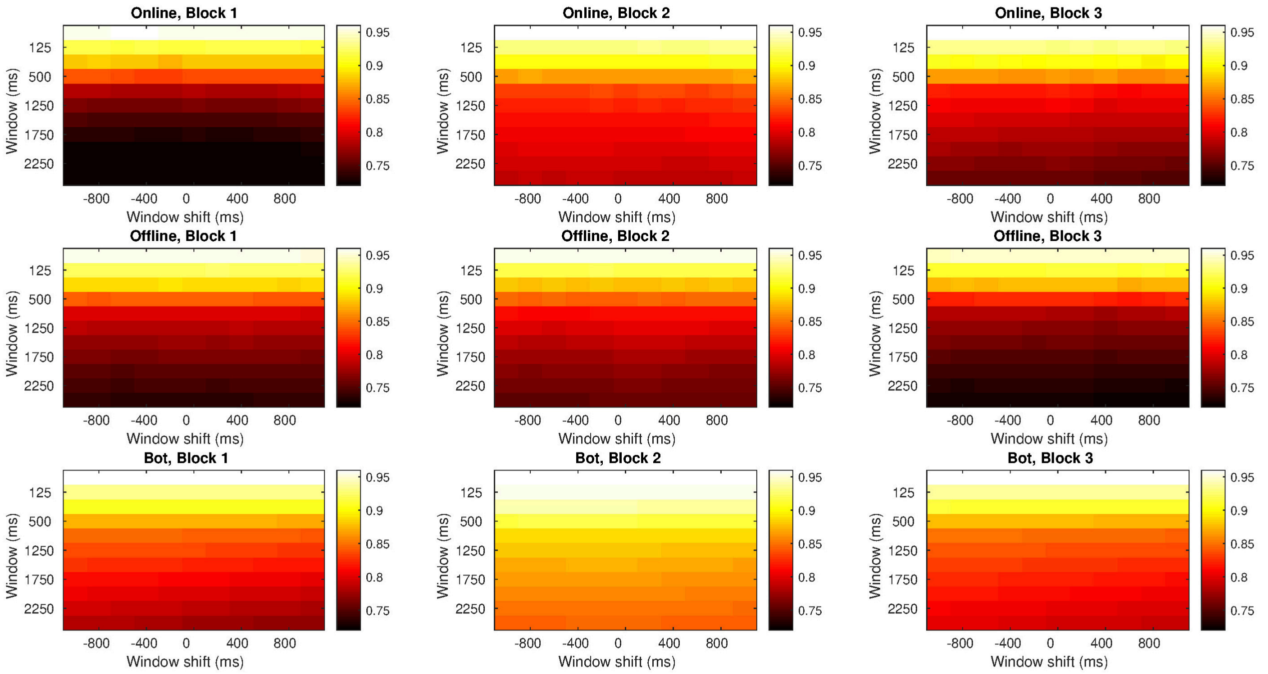


FIGURE 5 | Correlation indices for different windows span and delay times in Study 1. Each row refers to a type of agent (human online, human offline and oscillatory bot) and each column refers to either block 1, block 2, or block 3. Inside each graph, the horizontal axis is the delay applied (from −1 s to 1 s), with the central column showing no delay at all. Vertical axis alludes to the window span, from 50 ms (at the top) to 2500 ms (at the bottom). Light colors indicate higher correlation and dark colors indicate lower correlation. All the graphs show the same scale of colors; then, colors can be compared among graphs.

Similarity Between Two Series

Figure 5 shows the similarity between participants' series. On the one hand, light colors appeared at the top of the diagrams indicating a high correlation in small window spans. The smaller the windows span, the shorter the series were, and more similar to each other. On the other hand, colors got darker from the top to the bottom, which means that similarity gradually decreased as the window span increased.

Comparing the three blocks, the greatest similarity occurred in block 2 when the avatars were visible. But when considering the type of agent, regardless of the block, there was a greater similarity between the participant's and the bot's series than with the other agents.

In bigger window spans, the similarity with the human online series decreased: we can see that the colors at the bottom in block 1 are darker (i.e., less similarity) than the colors at the bottom in block 3. In big windows span, however, the similarity with the human offline series decreased along the study: the colors at the bottom of block 1 are lighter (i.e., more similarity) than the colors at the bottom of block 3.

Since there was no difference across the distinct delays applied (colors are uniformly distributed for diverse time gaps), we can employ the values with no delay. It means that we took the correlation index for window span with no delay applied to the trajectories in the statistical model we will present in section "Statistical Analysis" (for a table with all the correlation indices see the **Supplementary Material**).

Figure 6 takes the correlation values between two trajectories with no delay applied and shows how these indices change when different window spans are considered. On the one hand, **Figure 6A** shows that the highest correlations were found with the bot both in blocks 1 and 3. While in block 1 the correlation with the online agent was inferior than the correlation with the offline agent, in block 3 correlations with online were higher than the offline. On the other hand, **Figure 6B** indicates that the correlation values were higher in block 3 than in block 1 for the human online in all temporal scales. A similar pattern emerged with the oscillatory bot, with clearly higher correlations in block 3 in windows spans bigger than 1.5 s. The offline agent presented the opposite pattern since correlations decreased in block 3 in comparison to block 1.

Statistical Analysis

We first analyzed all the rounds. Many of the predictors of our model (age, gender, β and density of crossings) did not discriminate when the participant gave the correct answer (all p-Values > 0.05). Correlation indices between series with windows at 500 and 1200 ms seemed to discriminate participants' correct answer [Corr500, $t(1815) = -2.14$, $p = 0.03$; Corr1200, $t(1815) = 1.8$, $p = 0.06$]. However, the p-Value for the correlations for the second window span was not significant.

Secondly, we analyzed the results according to the type of agent and each block. Results, however, did not show any significant variable that predicts the correct response (all $ps > 0.05$) (see **Supplementary Material** for the detailed results).

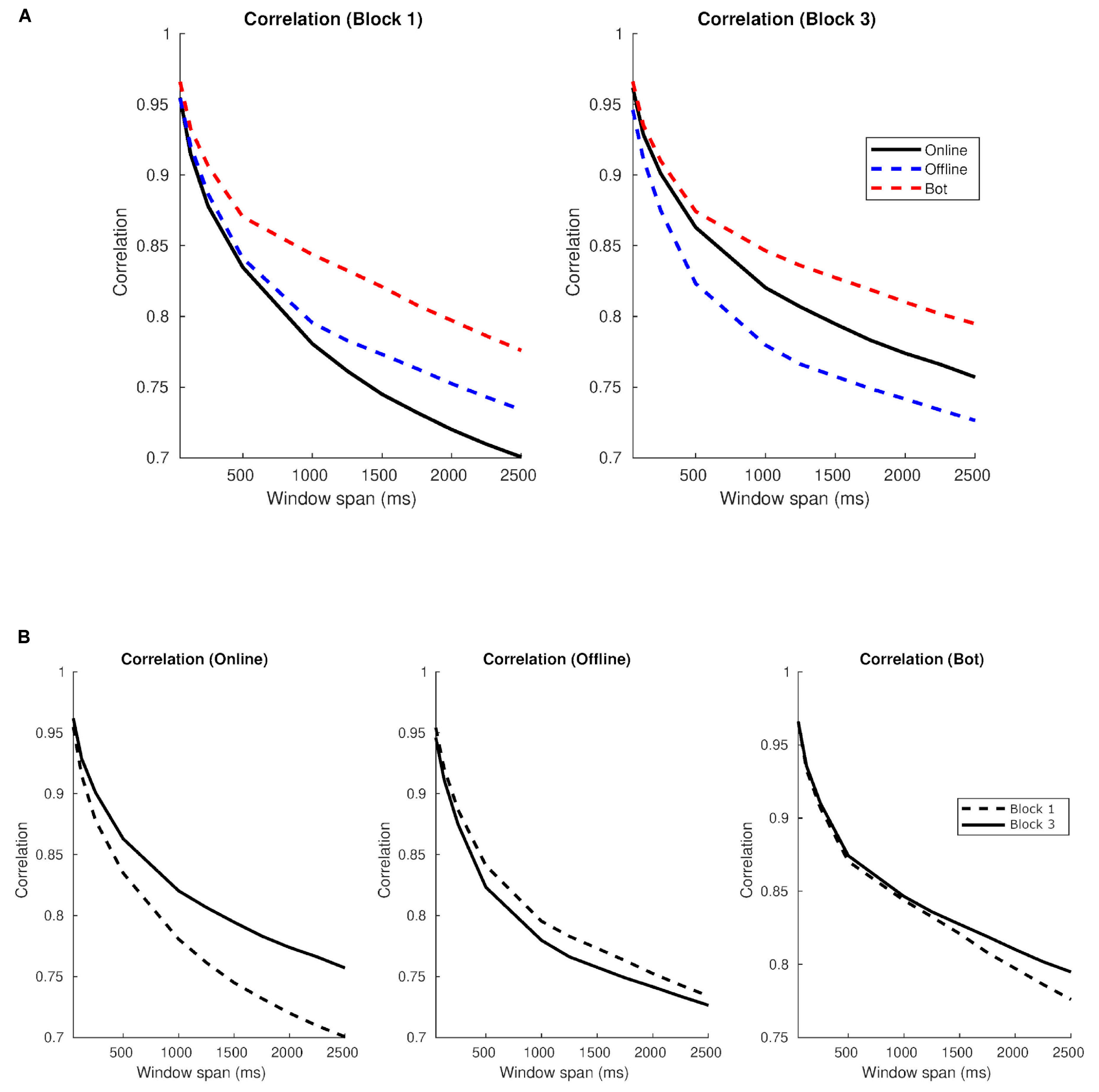


FIGURE 6 | Correlations in auditory blocks and per type of agent in Study 1. **(A)** Correlation indices in block 1 (left) and block 3 (right) per type of agent as a function of the window span. **(B)** Correlation indices per online agent (left), offline agent (medium) and oscillatory bot (right) in block 1 and block 3.

Debriefing

After the procedure was over, participants answered three questions:

- Describe briefly how you played the game.
- How have you decided that the agent was a human?
- How have you decided that the agent was a machine?

We grouped the responses into three categories:

(1) Reciprocity-based decisions: the participant explicitly said to ground her response on the reciprocity of movements, and/or the contingency between her movements and the other player's movements to decide her reply. They said, for example, that they searched the partner and then waited to see whether it reciprocated.
(2) Partially reciprocal decisions: the participant said to take reciprocity into account sometimes, but not always, and/or not in a consistent way. For instance, one said that she judged the agent was human because it responded to the participant's actions and because it moved randomly.
(3) Non-reciprocal responses: the participant said her decision was based on other features related to the pattern of sounds, the total number of crossings during a trial or the velocity of the movements.

In this study, only 11% of the participants' responses corresponded to category 1, 16% to category 2, and 73% to category 3.

Discussion

In our first study, participants explicitly differentiated the interaction with the human agents –online and offline– from the interaction with the oscillatory agent. Although they correctly identified the oscillatory bot as a machine, they treated the online and the offline agents as two instances of the same type of agent. When they interacted with either of them, they explicitly answered that they had interacted with a person. In a way, they were right, as the offline sequence was the replay of the previous online movements of a person. This suggests that responses were based on an observational stance rather than on the detection of the social contingencies available. Thus, the decisions took into account preferentially the degree of complexity of the sequence, responding "machine" when it exhibited a regular and predictable patters, and "person" when it did not. This hypothesis is reinforced by the fact that they performed alike across the three blocks. The audiovisual feedback of block 2 made no difference.

Interestingly, the different measures of an implicit interaction detection did not show that participants moved differently in the online vs. offline human conditions. Neither the features of the trajectories, the amount of crossings, nor the fractal analysis revealed a difference between online vs. offline agents across the blocks. Although the β for the interaction with the offline agent is similar to the index found in Bedia (a value that features uncorrelated processes), comparable indices were obtained for the other agents, again suggesting that our participants were trying to tell the nature of the other agent on the grounds of its global temporal pattern of sounds rather than on its contingency upon the participant moves.

On the other hand, the similarity between the two trajectories was greater in block 2 than in the other blocks, suggesting that the visual information available increased the matching of movements to those observed, but not enough to influence participants' response for that block. We found a greater correlation between the trajectories of agents at the 500 ms window span, suggesting that it is around this half second that contingencies may generate the experience of interaction, even if the participants' responses were not based on such experience, but on a periodical trajectory of the other agent.

In summary, participants' explicit responses in our first study were not based on the perceived contingency of the interaction, but presumably on the pattern of the other agent's movements ("if it is predictable and periodic, it is a machine; if not, it is a human"). As a matter of fact, the evidence indicates that they were not even able to implicitly discriminate between the online and offline conditions –in contraposition to previous results with the perceptual crossing paradigm. The reason for this, we submit, is that the offline condition can generate the experience of interaction, as the participant may respond contingently on the crossings detected. In addition, the fact that we provided two response options, "person" or "machine," pragmatically suggested to our participants that there were just two kinds of agents, inducing them to adopt an observational strategy over an interactive one; while, in fact, we presented three different kinds of moving agents in the set up. Similarly, the feedback provided after each trial, with so many errors, may have fostered the adoption of the "periodic pattern = machine, non-periodic = human" heuristic. The only possible way for participants to judge the offline cases as "machine" ones was by noticing their lack of responsivity to the participant moves, but either they did not look at this kind of information at all or the information available was not enough to detect such contingencies.

STUDY 2

In order to explore this *post hoc* hypothesis, we designed a second study. To better address the participants' attention to the pattern of interaction, we introduced modifications to the experimental paradigm. First, we only employed two types of agents: the human online and the human offline. In this way, given that the two kind of trajectories the participant could met were equivalent in complexity and unpredictability from an observer point of view, the only way to respond correctly would be by looking for reciprocal contingencies: moving and checking whether the other agent's movement was reactive to one's own movement. In particular, whether the pattern of back and forth movements of both agents around the same point emerged, and whether the participants could use this interaction pattern to respond to this new version of the "Turing test."

Secondly, we modified the way of giving the interrogation at the end of each round. Instead of choosing between two responses (person or machine), participants will decide the nature of the agent they have just interacted by selecting a point on a Likert scale. In this way, confidence in the response could be measured. No feedback about their answer was provided to prevent the development of strategies through trial and error along the procedure.

Materials and Methods

Participants

We tested 50 participants, recruited from a Psychology class, at the University of the Balearic Islands. Participants received credit points for participating in the study. In the lab, they were arranged in groups of 4 people (12 groups) and only one group of 2[3]. Two participants were excluded from the analysis because they had already participated in the first experiment. Our sample, then, consisted of 48 participants (9 males). Their ages ranged from 20 to 40 years (*M* = 23.45 and *SD* = 4.93).

Experimental Procedure

The procedure was similar to Study 1. After each participant entered into the cubicle and wore the headphones, 4 training trials were administered. During the training phase, the participant could see his/her avatar and his/her partner's avatar on a line on the screen. Each trial lasted for 15 s. At the end of each

---

[3] Although our lab counts with 6 cubicles, one computer was broken. Then, we arranged the experiment in groups of 4 people considering that the setup is programed to work with an even number of participants.

round, they had to detect whether they interacted with a human online or with a human offline. They replied by selecting one point on a 7-point Likert scale which represented how sure they were about the nature of the other agent. For instance, if they were completely sure that they had interacted with the human online, they chose one extreme of the scale that represented the response "I am completely sure it was online." If they were completely sure that they had interacted with the human offline, then they chose the other extreme of the scale that depicted the opposite reply; that is, "I am completely sure it was offline." In case they did not know the nature of the other agent, they chose the middle of the scale. Participants could also choose other two possibilities between the middle and each extreme of the scale. They did not receive feedback after each trial.

Each participant was tested along 6 rounds in block 1 (in which auditory feedback was provided), 6 rounds in block 2 (with audiovisual feedback) and 6 rounds in block 3 (again, only supplying auditory feedback). In each block, participants interacted with either a human online or a human offline agent in rounds that lasted 30 s. Like in the training phase, they had to decide, at the end of each round, whether they interacted with a human online or offline.

Coding

Like in Study 1, we coded participants' responses, the total number of crossings, fractal indices, the time between two crossings, the window of crossings, as well as the similarity between each pair of trajectories (the participant and the corresponding counterpart). In this case, participants' responses were considered as correct ones when they replied "online" after interacting with the human online agent, and "offline" in the rounds when they interacted with the human offline agent.

Statistical Approach

We again computed linear mixed effect models with the following predictors: type of agent (human online or human offline); type of block (block 1, block 2 and block 3), participant's age, gender, β, crossings, and correlation (similarity) between each pair of trajectories.

Results

Participants' Responses and Number of Crossings

In block 2, participants correctly identified the online agent (binomial two-sided test, $p < 0.001$). For the offline agent, although the probability of success was 0.53, it was not different from chance level (binomial two-sided test, $p = 0.5$). In all other blocks, participants' probability of success was not different from chance level for any type of agent (binomial two-sided tests, all ps > 0.56) and there was no difference in performance between block 1 and block 3 (see **Table 2**).

Considering only the online agent, the probability of success increased in block 2 (when participants could see the avatars on the screen) compared to block 1 and it decreased again in block 3. For the offline agent, the recognition rate was at chance level (between 0.49 and 0.53) in all blocks.

On the other hand, **Table 2** also shows the mean number of crossings that participants performed per block. In general, participants produced more crossings in block 1, for both types of agents, than in the rest of the blocks and they caused the fewest crossings in block 2. Participants collided more with the online than with the offline agent in both auditory blocks. In block 2, however, the mean number of crossings was similar for both agents. The probability of success and crossings are shown in **Figure 7**.

TABLE 2 | Probability of success and total number of crossings in Study 2.

	BLOCK 1	BLOCK 2	BLOCK 3
Correct answers			
Online	0.5	0.73	0.47
Offline	0.49	0.53	0.49
Number of crossings			
Online	42.92 (25.85)	25.65 (21.25)	36.99 (27.10)
Offline	38.88 (29.12)	27.59 (25.17)	30.57 (22.70)

Probability of correct responses about the nature of the other agent and mean (and standard deviation) of crossings in each block and type of agent in Study 2.

Fractal Indices

Figure 8 shows a graph with the distribution of the values of β per agent and block. Values of β for human-human online rounds were again lower than the values reported in Bedia et al. (2014; β = 0.86), while the values of β for the human offline agent were similar to the values of the shadow in Bedia et al. (2014; β = 0.29). Once more, β were practically identical in online and offline rounds and closer to 0 which characterizes uncorrelated processes (white noise; see **Figure 8**, bottom). For a table with all the values of β per agent and block, see **Supplementary Material**.

Time Between Two Crossings and Window of Crossings

Similar to Study 1, most crossings took place between 0.1 and 1 s after the previous crossing, as the center of the Gaussian distribution is between 10^2 and 10^3 ms (see **Figure 9A**, top). Also, most of the crossings from a round –after one crossing of reference– were distributed in a window span of 1 s (after the initial crossing), generating "gusts of crossings" (**Figure 9A**, bottom). In the time interval from 1 to 10 s, the number of accumulated crossings was lower (the sigmoid curve reached a plateau).

In blocks 1 and 3, more crossings occurred in the online agent condition than in the offline one, for the time interval lesser than 1 s. In block 2, the accumulated crossings –after one of reference– sharply decreased for both kinds of agents, in comparison to block 1, suggesting a greater role for the visual information.

Figure 9B shows the distribution of crossings as we increase the window span, between 500 and 1500 ms. In general, the number of crossings remained similar for the online agent from block 1 to block 3. In both blocks, when a participant crossed over the online agent, the following crossings summed up to four within the 1500 ms window span. This pattern was very similar for the offline agent in block 1 while it was slightly inferior in block 3: the offline reached at about 3 crossings, after one occurred, during the 1500 ms window span.

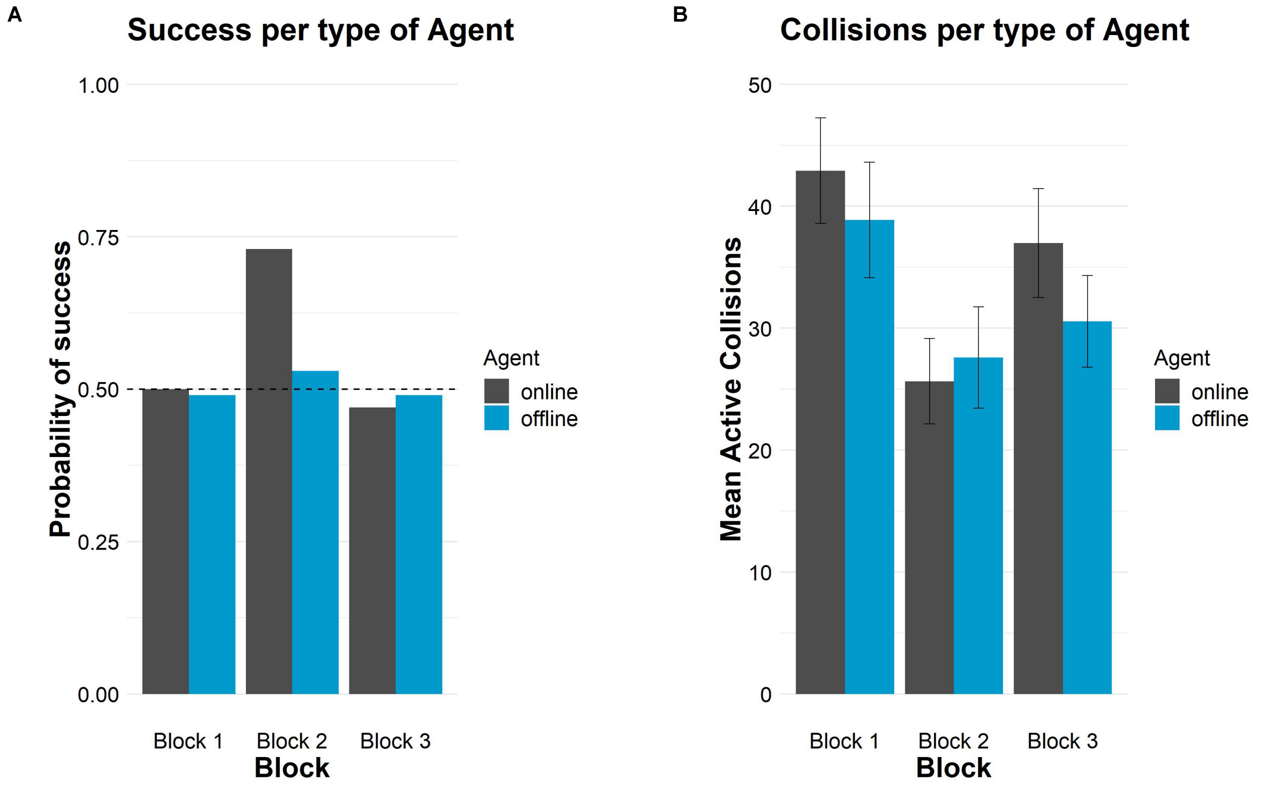


FIGURE 7 | Successes and crossings per type of agent and block in Study 2. **(A)** Probability of success in each block per type of agent. The horizontal dashed line represents chance level (50%). **(B)** Mean number of crossings in each block per type of agent. Error bars depict 95% confidence intervals.

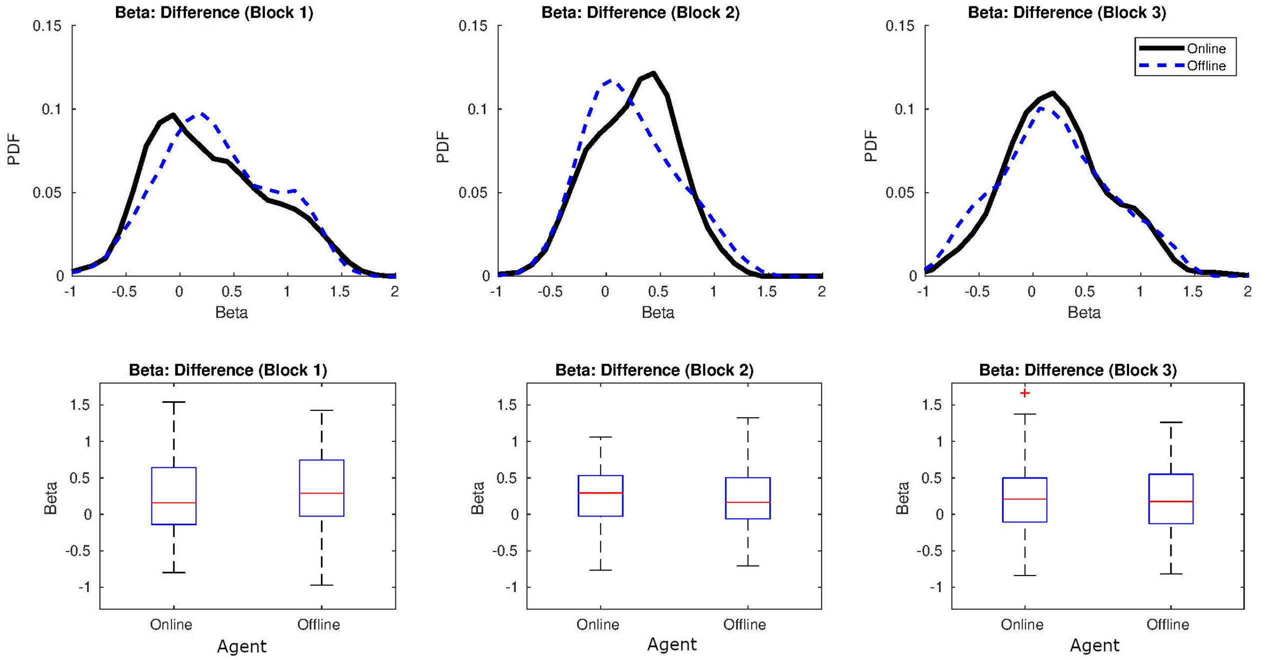


FIGURE 8 | Distribution of β in Study 2. At the top: distributions for the values of β (in black online agent, in blue offline agent). At the bottom, representation of the means and variances as boxplots.

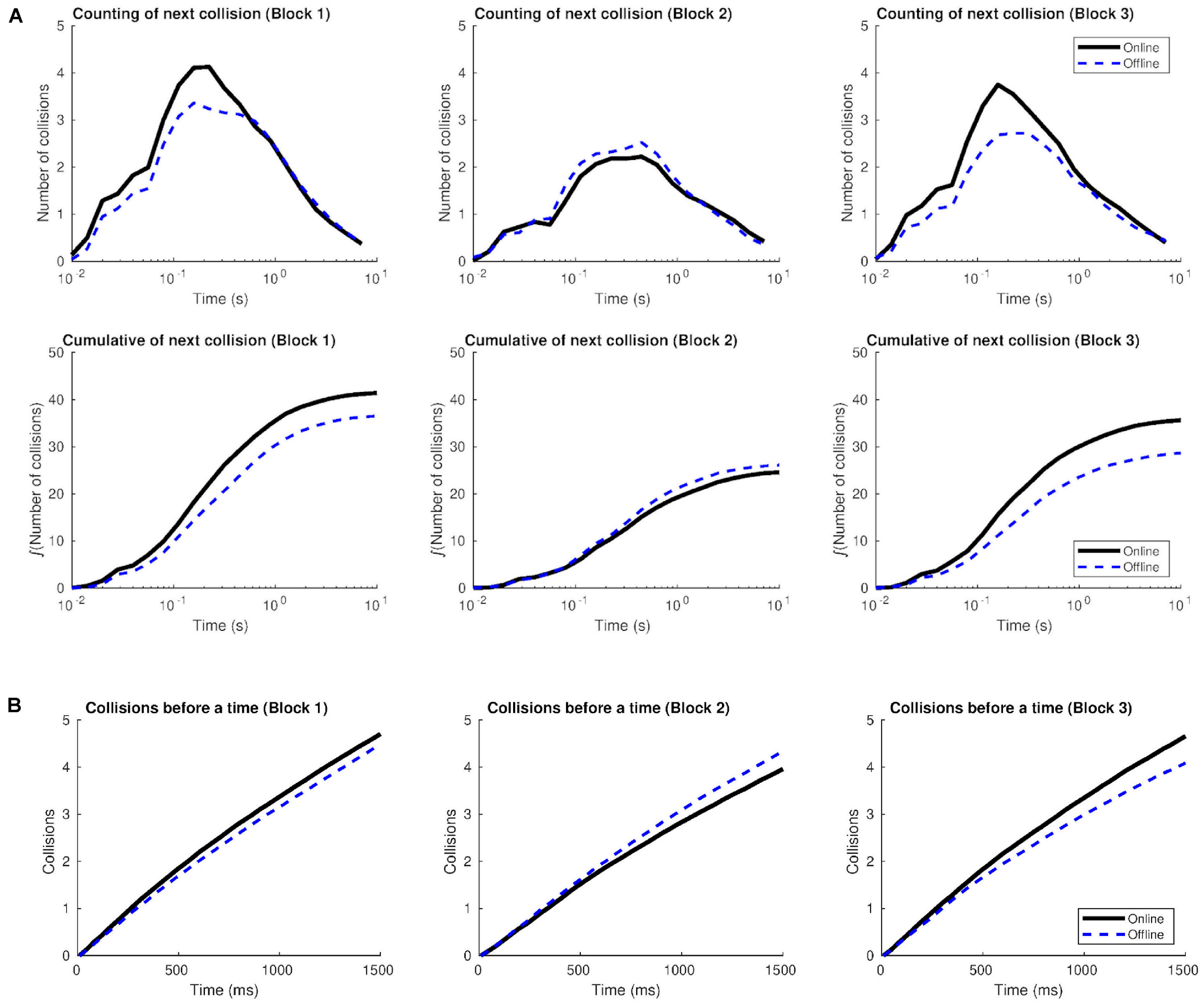


FIGURE 9 | Distribution of the number of crossings according to the window span in Study 2. **(A)** The distribution has a Gaussian shape. The *x*-axis (time) is shown on a logarithmic scale in order to see the relation between short times (milliseconds) and longer ones (dozens of seconds, which is the length of a round). At the top: number of crossings after one crossing of reference in relation to the time they are produced, for each block and type of agent. At the bottom: the number of accumulated crossings as a function of time. **(B)** Distribution of the number of crossings as we increase the window span. The graphs show how many crossings are produced in the window span from one crossing until a specific amount of time.

Similarity Between Two Series

Like in Study 1, correlations between series were very high when considering small time windows (in other words, the series of both interacting agents maximally resemble each other). This similarity slowly decreased as bigger time window were considered (see **Figure 10**). Also, like in Study 1, the greatest similarity occurred in block 2 when the other agent's movements were visible to the participant.

Colors were similarly distributed in blocks 1 and 3, indicating that correlation between trajectories were similar in both blocks, regardless of the type of agent involved. This suggests that, after the increase in correlation that occurred in the second block, the indices returned to the levels of block 1.

Since there was no difference across the distinct delays applied (colors were again uniformly distributed for diverse time delays), we employed the values with no delay. It means that we took the correlation index for the window span with no delay applied to the trajectories in the statistical model we will present in section "Statistical Analysis."

Figure 11 also takes the correlation values between two trajectories with no delay and represents how these indices change when different window spans are considered. On the one hand, **Figure 11A** shows that, while in block 1 the correlation with the online agent was slightly lower than the correlation with the offline agent, in block 3 the correlation indices were similar for both types of agents. On the other hand, **Figure 11B** indicates that correlations were higher in block 1 than in block 3 for both online and offline agents in all window spans.

Statistical Analysis

We built a model containing several predictors. We found, on the one hand, that block 2 discriminated participants' correct

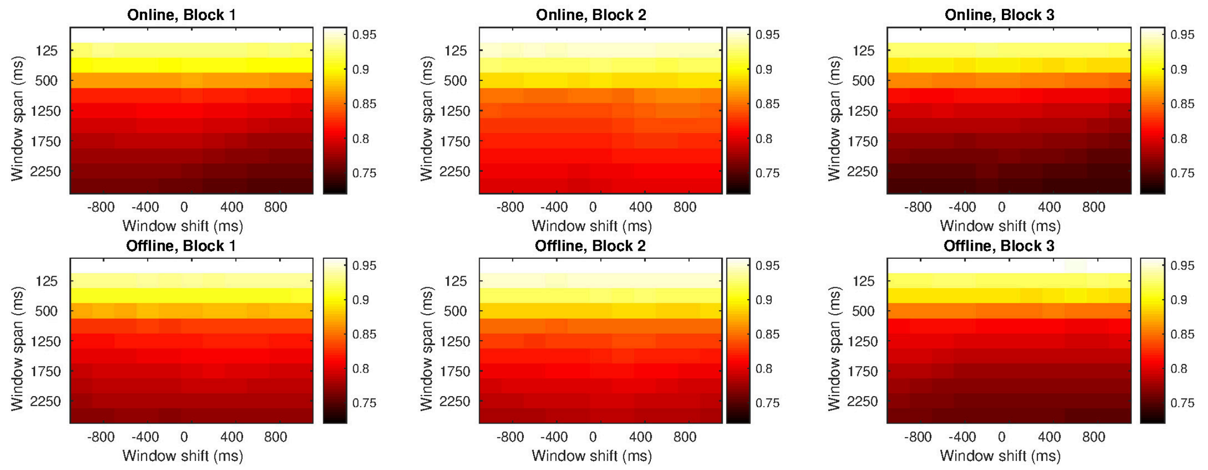


FIGURE 10 | Correlation indices for different windows span and delay times in Study 2. Each row refers to a type of agent (human online and human offline) and each column refers to either block 1, block 2, or block 3. Inside each graph, the horizontal axis is the delay applied (from –1 s to 1 s), with the central column showing no delay at all. Vertical axis alludes to the window span and it ranges from 50 ms **(at the top)** to 2500 ms **(at the bottom)**. Light colors indicate higher correlation and dark colors indicate lower correlation. All the graphs show the same scale of colors; then, colors can be compared among graphs.

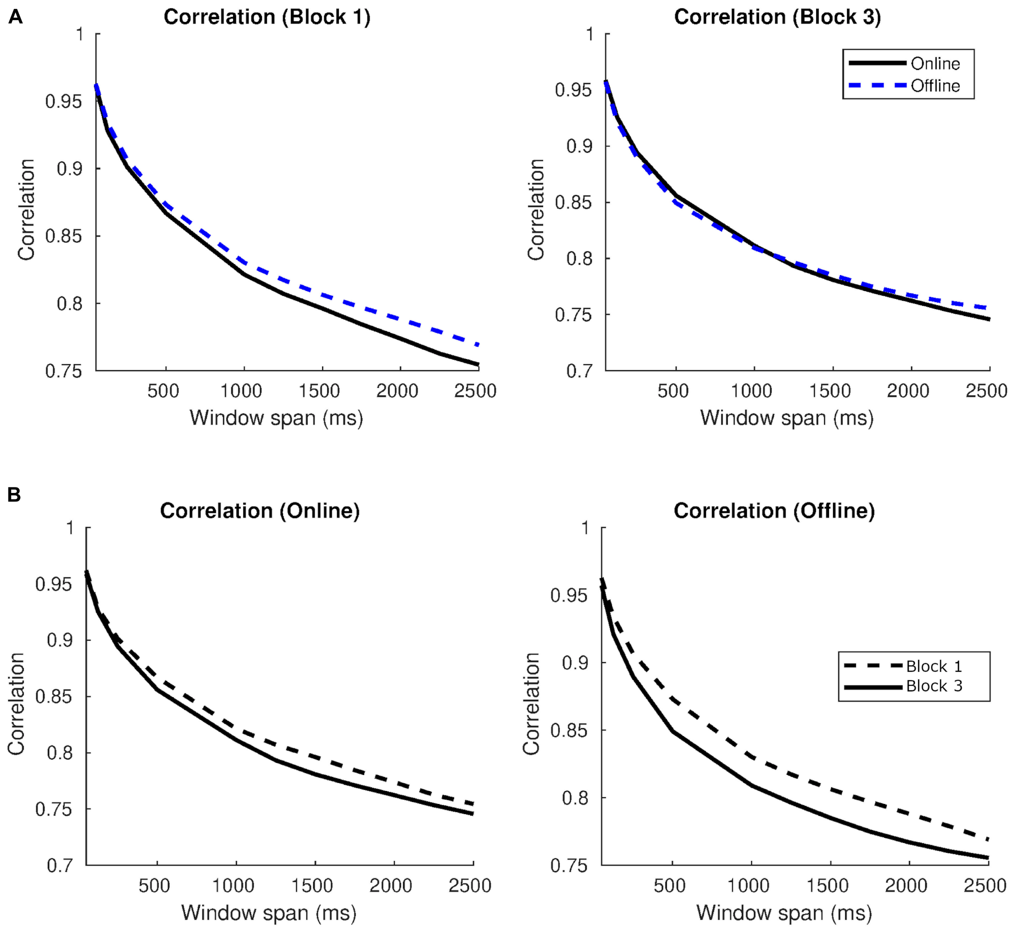


FIGURE 11 | Correlations in auditory blocks and per type of agent in Study 2. **(A)** Correlation indices in block 1 (left) and block 3 (right) per type of agent as a function of the window span. **(B)** Correlation indices per online agent (left) and offline agent (right) in block 1 and block 3.

TABLE 3 | Probability of success and total number of crossings.

	BLOCK 1	BLOCK 2	BLOCK 3
Correct answers			
Online	0.6	0.8	0.59
Offline	0.51	0.63	0.53
Number of crossings			
Online	47.28 (27.64)	27.38 (13.99)	39.78 (30.28)
Offline	45.86 (51.14)	27.59 (29.58)	30.10 (25.60)

Probability of correct responses about the nature of the other agent and mean (and standard deviation) of crossings in each block and type of agent.

answer [$t(837) = 3.53$, $p < 0.001$]. On the other hand, all the other predictors (type of agent, age, gender, beta, density of crossings and correlation indices) did not discriminate when the participant gave the correct answer (ps > 0.05; see **Supplementary Material** for the detailed results).

These results differ from what we found in Study 1: now the audiovisual information provided in block 2 allowed participants to correctly distinguish both kinds of agents.

Debriefing

As in Study 1, at the end of the procedure we asked the same three questions about their experience in the test to the participants:

- Describe briefly how you played the game.
- How have you decided that the agent was online?
- How have you decided that the agent was offline?

This time, 65% of participants' responses corresponded to category 1 ("reciprocity-based" responses), 25% to category 2 ("partially reciprocity-based" responses), and 10% to category 3 ("non-reciprocity based" responses). We take this as evidence that our second study managed to induce participants to rely on interaction detection, even if the information available was not discriminant enough, except for the audiovisual block.

Confidence in the Response

We also analyzed the performance in the rounds that participants reported to feel completely sure of their response. We took into account the extreme points of the 7-point Likert scale that represented the responses "I am completely sure that I interacted with the Online agent" (1) and "I am completely sure that I interacted with the Offline agent" (7) and selected the rounds in which participants gave that reply.

A total of 303 series, out of 860, qualified as maximally confident ones. Most participants (45 out of 48) were represented in this subset. In those trials, the participants correctly identified the online agent (binomial two-sided test, $p < 0.001$) and the offline agent (binomial two-sided test, $p = 0.05$) in block 2. Their probability of success was not different from chance level for any type of agent in blocks 1 and 3 (binomial two-sided tests, all ps > 0.23). In general, the performance was different from chance level only in block 2, although it was always over 0.5 in all blocks (see **Table 3**).

Regarding crossings, self-confident participants produced more crossings in block 1, in the two conditions, than in the rest of the blocks. As in Study 1, they produced, on average, fewer crossings in block 2 than in the rest of the blocks (see also **Table 3**). In general, there were more crossings with the online than the offline agent in blocks 1 and 3. In block 2 there were the same number of crossings for both types of agents. **Figure 12** shows the probability of success and the crossings for online and offline agents in the three blocks.

The type of block had a significant effect on crossings, $\chi^2(7) = 16.39$, $p < 0.001$, but not the type of agent, $\chi^2(8) = 0.50$, $p = 0.47$. Most important, the block × agent interaction was not significant, $\chi^2(10) = 1.46$, $p = 0.48$.

Discussion

The changes to our experimental paradigm introduced in our second study were able to reveal the use of reciprocal contingencies to detect an interaction and to explicitly respond the question task, but only when the available information was sufficient to do so: participants were able to correctly distinguish whether they were interacting with another person when they could see the other moving as they moved, but not when just hearing a sound when a crossing took place. To put it another way, participants responded at chance level when they could not tell apart whether an interaction was taking place or not, as it happened in the auditory condition. The perceptual crossing paradigm, which only provides acoustic (or tactile) feedback of the crossing, does not seem to provide enough discriminant information to detect social contingencies.

This insufficiency is made clearer when it is realized that the participant can adopt either an active or a passive strategy in addressing the task. He/she may move along the axis trying to cross over the other agent, to see what happens next (i.e., active strategy), or he/she may wait for a crossing over to happen, to see whether the other agent moved back, generating the characteristic series of crossings around a point (i.e., passive strategy). Moreover, they may combine both strategies in a single trial. But given that the offline agent in the second study was a trajectory that had already been correctly deemed as human, they might have generated an illusion of interaction, as these contingencies were in fact available in the participants –except in the audiovisual condition, where the participants correctly classified the agent, on the unique evidence available, the interactive one. In this way, Study 2 also confirms our hypothesis that interaction detection involves the detection of contingencies within a short window span, inferior to 1 s –basically around 500 ms, as found in Study 1.

GENERAL DISCUSSION

In this article, we have shown that individuals can resort to social contingencies to respond in a minimal Turing test. Based on a new version of the perceptual crossing paradigm, we also investigated in two studies how much information is required to reliably detect the reciprocal sensorimotor contingencies in social interactions and, therefore, recognize the nature of the other agent.

In particular, we presented participants with movements of a human agent, either online or offline, and movements of

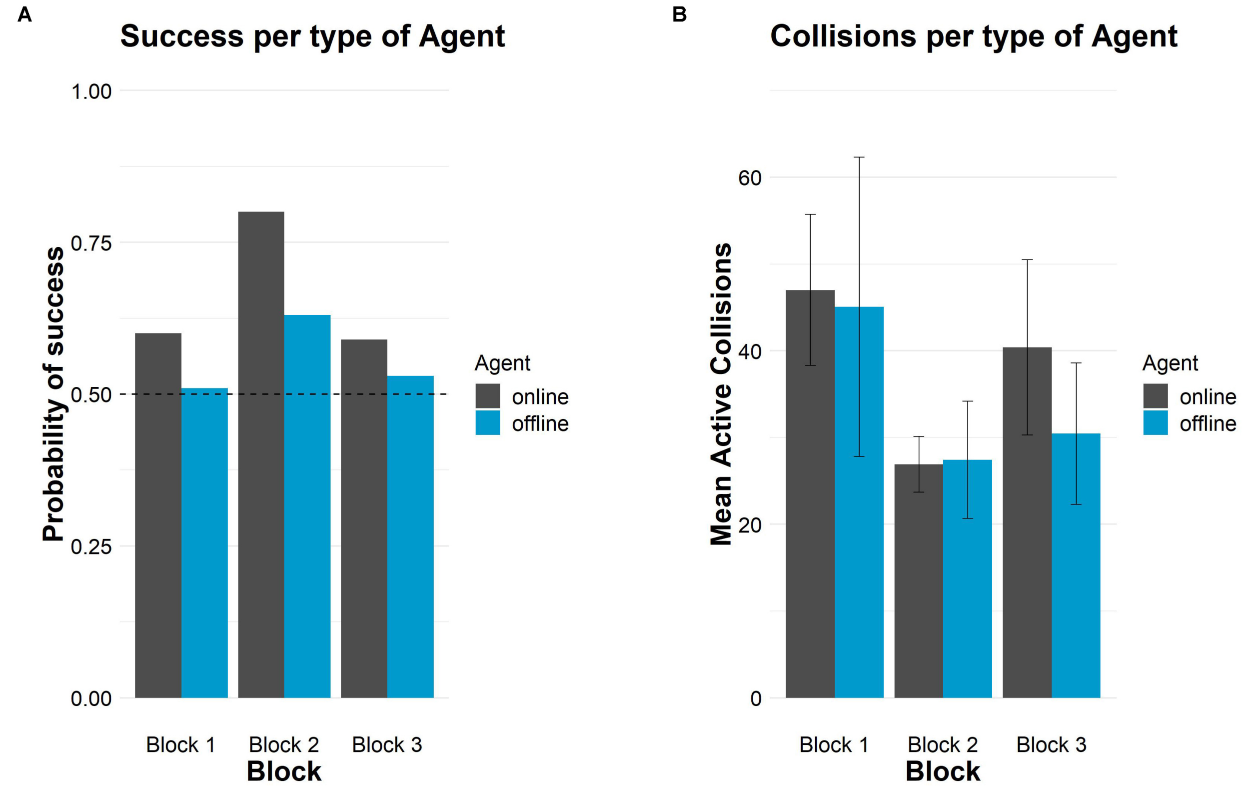


FIGURE 12 | Successes and crossings per type of agent and block for the self-confident sample. **(A)** Probability of success in each block per type of agent. The horizontal dashed line represents chance level (50%). **(B)** Mean number of crossings in each block per type of agent. Error bars depict 95% confidence intervals.

an oscillatory agent in three different blocks. In each block, participants received either auditory or audiovisual feedback along each trial. Analysis of participants' explicit responses and of the implicit information subsumed in the dynamics of their series revealed that participants use the reciprocal sensorimotor contingencies of the interaction in very specific scenarios.

The results of Study 1 showed that although participants differentiated the human agents from the oscillatory agent and correctly identified the latter one as a machine, they judged both the online and the offline agents as persons, and their performance did not improve in the audiovisual condition. Analysis of the implicit measures during the interaction revealed no specific pattern for each type of agent. According to these results, participants may have used the periodicity of the other's movements to base their response instead of detecting the interaction. The implicit measures of interaction agree with the participant's reports at the end of the procedure.

The instructions given in Study 1 failed to induce the participants to look for the cues of interaction: while they expected to interact either with a human or a machine, we presented them with three different traces that corresponded to three types of agents (online, offline, and the oscillatory bot). In order to better induce the participant to pay attention to the interaction cues, in Study 2 we eliminated the oscillatory bot and changed the question to tell whether the other agents were online or offline. We also adjusted the way participants replied at the end of each round: they had to select, on a 7-point Likert scale, how sure they were if the other agent was online or offline. The results showed that participants correctly identified the online agent in the audiovisual condition but they failed to recognize the offline agent in all blocks, even when they could see the avatars on the screen.

Only when we analyze the performance of the participants who were completely sure of their reply, correct recognition of both the online and the offline agents in the audiovisual condition emerged. Therefore, our results show that participants can base their responses on the coupled dynamics of interaction. However, auditory signaling the crossing is not enough to tell apart real interaction from a previously recorded, rightly recognized as interactive, trajectories. This is congruent with previous studies in the perceptual crossing framework. They relied on one modality only (tactile stimulation or auditory feedback) and kept that stimulation constant along the study and found the difficulty in discriminating the avatar of the human agent from the mobile lure. Associating different tones to each type of agent may be more informative (Lenay and Stewart, 2012) but might also induce participants to pay attention to the sounds themselves instead to the pattern of the interaction.

Therefore, our study shows that participants are able to use reciprocal sensorimotor contingencies of the interaction, even if the minimal information available may require more than one sensory modality. As a matter of fact, only when both auditory and visual cues are provided participants became reliable in distinguishing the type of agent they are interacting with. This finding is congruent with the intersensory redundancy hypothesis (Bahrick and Lickliter, 2012) according to which amodal properties are best perceived when simultaneous co-occurrence of stimulation across different sense modalities takes place. In our task, the temporal synchrony of the interaction was detected in the audiovisual scenario when participants simultaneously perceived visual and auditory information of the interaction. Such intersensory facilitation happens because "redundantly specified, amodal properties are highly salient and [thus] detected more easily in bimodal synchronous stimulation than [. . .] the same amodal properties in unimodal stimulation" (Bahrick and Lickliter, 2012, 188).

Interestingly, we did not find a significant difference between the first and the third block in any of the studies. The audiovisual block 2 did not have an effect in the last block, again suggesting the insufficiency of unimodal information for interaction detection.

Our studies also confirmed the utility of the implicit measures of interaction introduced by Bedia et al. (2014). The correlation between the series of two players, showed that the series resembled more to each other in the audiovisual condition at the 500 ms window span. This means that the participants tended to assimilate their trajectories to those of the agents they were interacting with when they could see them. It is also around this half-second that contingencies may generate the experience of interaction. Interestingly, developmental studies revealed that social contingency between different response modalities of infants and their mothers can also be appreciated in a similar window span equal to or less than 1 s long (Dominguez et al., 2016; Español et al., under revision). On the one hand, newborns' and maternal vocalizations occurred within a 1-s window (Dominguez et al., 2016) and, on the other hand, maternal responsiveness through patterns of imitation and affect attunement to the infant's signals during the first 10 months also unfold during these short time windows (Español et al., under revision). In general, the window span for cycles of reciprocity using different response patterns does not seem to vary through development.

However, fractal analyses did not result in any specific pattern per type of agent. This null result may be partially explained by the audiovisual condition. In the original study of Bedia et al. (2014) "the emergence of a 1/*f* structure for genuine social interaction is something that happens only in the shared space between the two subjects, and the process cannot be reduced to the individual dynamics of any of them" (p. 11). In our study, the shared space was not maintained exactly constant along the whole experiment since the feedback was different in each block. This difference seems to be critical for this measure.

Finally, the introduction of the human offline agent may have caused some difficulties in the procedure. If we compare it with the previous shadow and lure bots from other perceptual crossing experiments, the behavior of our offline agent was more complex: it was able to create the illusion of interaction because it was originally the outcome of a previous contingent interaction between two humans. This additional complexity may have confused the participants, making it very difficult to discriminate between the online and the offline cases and struggling also the way they interacted with each agent. Although previous research showed that the detection of the offline agent is possible in a tactile setup with pair of participants (Iizuka et al., 2009), this achievement required tens of trials. We did not run the study along such many trials to test whether recognition finally emerged over extended periods of interactions, but we think that it possibly does.

There is an additional limitation in our study: although a participant interacted with one agent at a time, the behavior of the human online agent did not correspond to the *same* human participant along the trials. As we performed a group experiment (groups of 4 or 6 participants in a row), each trial in which a participant interacted with a human online was randomly assigned to a group, so he/she never interacted with the same counterpart in the online condition. That is, each participant encountered up to 5 different people in Study 1, or up to 3 human counterparts in Study 2. Previous experiments studied how the detection of agency emerged in pair of participants as the other human of this pair was the same person during the whole setup (except for Bedia et al., 2014). As each person can display different strategies and behaviors, this source of variation could increase the difficulty of the task. More research is needed to delve into each of these issues.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any researcher.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of ethical guidelines of the Research Ethics Committee of the University of the Balearic Islands. All participants gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Research Ethics Committee of the University of the Balearic Islands.

AUTHOR CONTRIBUTIONS

PB and AG conceived of the presented idea, designed the study, and wrote the manuscript. MB worked out almost all of the technical details of the analyses. PB carried out the experiments. PB and MB performed the analyses and designed the figures and tables. AG supervised the project.

FUNDING

This work was supported by the research project FFI2017-86351-R of the Spanish Government (AEI/FEDER, EU). PB was supported by a Ph.D. scholarship from the Spanish Government (Grant No. BES-2014-067640). MB thanks the support of the project TIN2016-80347-R funded by the Spanish Ministry of Economy and Competitiveness.

ACKNOWLEDGMENTS

We thank the students who took part in both studies and the participants of the socSMCs conference 2018 "Social cognition in humans and robots" for their very valuable comments and questions on the poster version of this manuscript. We also thank David Gracia Larrodé and Tomás Gómez for their valuable help in this project, as well as the two referees from this journal for their questions and suggestions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2020.00102/full#supplementary-material

REFERENCES

Auvray, M., Lenay, C., and Stewart, J. (2009). Perceptual interactions in a minimalist virtual environment. *New Ideas Psychol.* 27, 32–47. doi: 10.1016/j.newideapsych.2007.12.002

Auvray, M., and Rohde, M. (2012). Perceptual crossing: the simplest online paradigm. *Front. Hum. Neurosci.* 6:181. doi: 10.3389/fnhum.2012.00181

Bahrick, L. E., and Lickliter, R. (2012). "The role of intersensory redundacy in early perceptual, cognitive, and social development," in *Multisensory Development*, eds A. J. Bremner, D. J. Lewkowicz, and C. Spence (Oxford: Oxford University Press), 183–205.

Bedia, M. G., Aguilera, M., Gómez, T., Larrode, D. G., and Seron, F. (2014). Quantifying long-range correlations and 1/f patterns in a minimal experiment of social interaction. *Front. Psychol.* 5:1281. doi: 10.3389/fpsyg.2014.01281

Berkovits, I., Hancock, G. R., and Nevitt, J. (2000). Bootstrap resampling approaches for repeated measure designs: relative robustness to sphericity and normality violations. *Educ. Psychol. Meas.* 60, 877–892. doi: 10.1177/00131640021970961

Bigelow, A. E., and Decoste, C. (2003). Sensitivity to social contingency from mothers and strangers in 2-, 4-, and 6-month-old infants. *Infancy* 4, 111–140. doi: 10.1207/S15327078IN0401_6

Bigelow, A. E., MacLean, B. K., and MacDonald, D. (1996). Infants' response to live and replay interactions with self and mother. *Merrill Palmer Q.* 42, 596–611. doi: 10.2307/23087472

Dominguez, S., Devouche, E., Apter, G., and Gratier, M. (2016). The roots of turn-taking in the neonatal period. *Infant Child Dev.* 25, 240–255. doi: 10.1002/icd.1976

Field, T., Nadel, J., Hernandez-reif, M., Diego, M., Vera, Y., Gil, K., et al. (2005). Depressed mothers' infants show less negative affect during non-contingent interactions. *Infant Behav. Dev.* 28, 426–430. doi: 10.1016/j.infbeh.2005.03.003

Gelman, A., Hill, J., and Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *J. Res. Educ. Eff.* 5, 189–211. doi: 10.1080/19345747.2011.618213

Gilden, D. L. (2001). Cognitive emissions of 1/f noise. *Psychol. Rev.* 108, 33–56. doi: 10.I037//0033-295X.108.1.33

Gomila, A. (2001). La perspectiva de segunda persona: mecanismos mentales de la intersubjetividad. *Contrastes* 6, 65–86. doi: 10.24310/Contrastescontrastes.v0i0.1448

Gomila, A. (2002). La perspectiva de segunda persona de la atribución mental. *Azafea Rev. Filos.* 4, 123–138.

Gomila, A. (2012). *Verbal Minds. Language and the Architecture of Cognition*. London: Elsevier.

Gomila, A., and Pérez, D. (2017). "Lo que la segunda persona no es," in *La Segunda Persona y Las Emociones*, eds D. Perez, and D. Lawler (Ciudad Autónoma de buenos Aires: SADAF), 275–297.

Iizuka, H., Ando, H., and Maeda, T. (2009). "The anticipation of human behavior using 'parasitic humanoid,'," in *Human-Computer Interaction. Ambient, Ubiquitous and Intelligent Interaction*, ed. J. Jacko (Berlin: Springer-Verlag), 284–293. doi: 10.1007/978-3-642-02580-8_31

Lenay, C., and Stewart, J. (2012). Minimalist approach to perceptual interactions. *Front. Hum. Neurosci.* 6:98. doi: 10.3389/fnhum.2012.00098

Lenay, C., Stewart, J., Rohde, M., and Amar, A. A. (2011). "You never fail to surprise me": the hallmark of the other: experimental study and simulations of perceptual crossing. *Interact. Stud.* 12, 373–396. doi: 10.1075/is.12.3.01len

Murray, L., and Trevarthen, C. (1985). "Emotional regulation of interactions between two-month-olds and their mothers," in *Social Perception in Infants*, eds T. Field, and N. Fox (Norwood, NJ: Ablex), 177–197.

Nadel, J., Carchon, I., Kervella, C., Marcelli, D., and Réserbat-Plantey, D. (1999). Expectancies for social contingency in 2-month-olds. *Dev. Sci.* 2, 164–173. doi: 10.1111/1467-7687.00065

Okanda, M., and Itakura, S. (2005). "Young infants' sensitivity to social contingency from mother and stranger: Developmental changes," in *Proceedings of the 4th International Conference on Development and Learning* (Osaka: IEEE), 165–168. doi: 10.1109/DEVLRN.2005.1490971

Peng, C. K., Hausdorff, J. M., and Goldberger, A. L. (2000). "Fractal mechanisms in neuronal control: human heartbeat and gait dynamics in health and disease," in *Self-Organized Biological Dynamics and Nonlinear Control*, ed. J. Walleczek (Cambridge: Cambridge University Press), 66–96. doi: 10.1017/cbo9780511535338.006

Pérez, D. (2013). *Sentir, Desear, Creer: Una Aproximación Filosófica a Los Conceptos Psicológicos*. Buenos Aires: Prometeo.

Pfeiffer, U. J., Timmermans, B., Bente, G., Vogeley, K., and Schilbach, L. (2011). A non-verbal turing test: differentiating mind from machine in gaze-based social interaction. *PLoS One* 6:e27591. doi: 10.1371/journal.pone.0027591

Reddy, V. (2008). *How Infants Know Minds*. Harvard: Harvard University Press.

Stormark, K. M., and Braarud, H. C. (2004). Infants' sensitivity to social contingency: a "double video" study of face-to-face communication between 2- and 4-month-olds and their mothers. *Infant Behav. Dev.* 27, 195–203. doi: 10.1016/j.infbeh.2003.09.004

Trevarthen, C. (1979). "Communication and cooperation in early infancy: a description of primary intersubjectivity," in *Before Speech: The Beginning of Human Communication*, ed. M. Bullowa (London: Cambridge University Press), 321–347.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind* 59, 433–460. doi: 10.1093/mind/LIX.236.433

Van Orden, G. C., Holden, J. G., and Turvey, M. T. (2003). Self-organization of cognitive performance. *J. Exp. Psychol. Gen.* 132, 331–350. doi: 10.1037/0096-3445.132.3.331

Van Orden, G. C., Holden, J. G., and Turvey, M. T. (2005). Human cognition and 1/f scaling. *J. Exp. Psychol. Gen.* 134, 117–123. doi: 10.1037/0096-3445.134.1.117

Westfall, P. H. (2011). On using the bootstrap for multiple comparisons. *J. Biopharm. Stat.* 21, 1187–1205. doi: 10.1080/10543406.2011.607751

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Robotics and AI

ORIGINAL RESEARCH
published: 16 April 2020
doi: 10.3389/frobt.2020.00047

Imitating by Generating: Deep Generative Models for Imitation of Interactive Tasks

Judith Bütepage[1]*, **Ali Ghadirzadeh**[1,2], **Özge Öztimur Karadağ**[1,3], **Mårten Björkman**[1] and **Danica Kragic**[1]

[1] Robotics, Perception and Learning, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden, [2] Intelligent Robotics Research Group, Aalto University, Espoo, Finland, [3] Department of Computer Engineering, Alanya Alaaddin Keykubat University, Antalya, Turkey

To coordinate actions with an interaction partner requires a constant exchange of sensorimotor signals. Humans acquire these skills in infancy and early childhood mostly by imitation learning and active engagement with a skilled partner. They require the ability to predict and adapt to one's partner during an interaction. In this work we want to explore these ideas in a human-robot interaction setting in which a robot is required to learn interactive tasks from a combination of observational and kinesthetic learning. To this end, we propose a deep learning framework consisting of a number of components for (1) human and robot motion embedding, (2) motion prediction of the human partner, and (3) generation of robot joint trajectories matching the human motion. As long-term motion prediction methods often suffer from the problem of regression to the mean, our technical contribution here is a novel probabilistic latent variable model which does not predict in joint space but in latent space. To test the proposed method, we collect human-human interaction data and human-robot interaction data of four interactive tasks "hand-shake," "hand-wave," "parachute fist-bump," and "rocket fist-bump." We demonstrate experimentally the importance of predictive and adaptive components as well as low-level abstractions to successfully learn to imitate human behavior in interactive social tasks.

Keywords: imitation learning, human-robot interaction, generative models, deep learning, sensorimotor coordination, variational autoencoders

Edited by:
Séverin Lemaignan,
Bristol Robotics Laboratory,
United Kingdom

Reviewed by:
Soheil Keshmiri,
Advanced Telecommunications
Research Institute International (ATR),
Japan
Chenguang Yang,
University of the West of England,
United Kingdom

****Correspondence:***
Judith Bütepage
butepage@kth.se

Specialty section:
This article was submitted to
Human-Robot Interaction,
a section of the journal
Frontiers in Robotics and AI

Received: *31 October 2019*
Accepted: *17 March 2020*
Published: *16 April 2020*

Citation:
Bütepage J, Ghadirzadeh A,
Öztimur Karadağ Ö, Björkman M and
Kragic D (2020) Imitating by
Generating: Deep Generative Models
for Imitation of Interactive Tasks.
Front. Robot. AI 7:47.
doi: 10.3389/frobt.2020.00047

1. INTRODUCTION

Physical human-robot interaction requires the robot to actively engage in joint action with human partners. In this work, we are interested in robotic learning of physical human-robot tasks which require coordinated actions. We take inspiration from psychological and biological research and investigate how observational and kinesthetic learning can be combined to learn specific coordinated actions, namely interactive greeting gestures.

In a more general context, coordinated actions between humans can be of functional nature, such as handing over an object, or of social importance, such as shaking hands as a greeting gesture. Thus, joint actions encompass any kind of coordination of actions in space and time in a social context. In general, joint actions require the ability to share representations, to predict others' actions and to integrate these predictions into action planning (Sebanz et al., 2006). On a sensorimotor level coordinated actions require a constant coupling between the partners' sensory and motor channels (Vesper et al., 2017). We aim at making use of sensorimotor patterns to enable a robot to engage with a human partner in actions that require a high degree of coordination such as hand-shaking.

The acquisition of the ability to engage in joint action during human infancy and early childhood is an active field of research in psychology (Brownell, 2011). Interaction is mostly learned in interaction, from observation, active participation, or explicit teaching. While cultural differences exist, children are commonly presented with the opportunity to learn through guided participation in joint action with more experienced interacting partners (Rogoff et al., 1993). In the robotics community two prominent techniques to learn actions from others are *learning from demonstration* and *imitation learning* (Billard et al., 2008; Osa et al., 2018). Learning from demonstration can be seen as a form of imitation learning. It requires a teacher to intentionally demonstrate to a learner how an action should be performed. In a robotic learning scenario, this can imply direct kinesthetic teaching of trajectories. General imitation learning on the other hand includes also learners who passively observe an action and replicate it without supervision. When observing a human, who often has a different set of degrees of freedom, the robotic system first needs to acquire a mapping between embodiments before a motion can be imitated (Alissandrakis et al., 2007).

In this work, we are interested in teaching a robot to coordinate with a human in time and space. Therefore, we require adaptive and predictive models of sensorimotor patterns such as joint trajectories and motor commands of interactive tasks. To this end, we develop deep generative models that represent joint distributions over all relevant variables over time. The temporal latent variables in these models encode the underlying dynamics of the task and allow for a sensorimortor coupling of the human and the robot partner. As depicted in **Figure 1**, collecting data by kinesthetic teaching for human-robot interaction tasks is tedious and time-consuming. We propose to first model the dynamics of human-human interaction and subsequently use the learned representation to guide the robot's action selection during human-robot interaction.

Before diving into the theory, in the next section we will shortly introduce the field of robotic imitation learning and point out how the general field differs from the requirements needed for imitation learning for interaction. Finally, we will motivate our choice of model and explain the basic assumptions of deep generative models.

2. BACKGROUND

Traditionally, robotic imitation learning is applied to individual tasks in which the robot has to acquire e.g., motor skills and models of the environment. Our goal is to extend these ideas to interactive settings in which a human partner has to participate in action selection. Thus, we aim at transferring knowledge form observing human-human interaction (HHI) to human-robot interaction (HRI).

2.1. Robotic Imitation Learning of Trajectories

Imitation learning is concerned with acquiring a policy, i.e., a function that generates the optimal action given an observed state. While reinforcement learning usually solves this task with help of active exploration by the learning agent, in imitation learning the agent is provided with observations of states and actions from which to learn. These demonstrations can either be generated in the agent's own state space, e.g., by tele-operation (Argall et al., 2009), or in the demonstrators embodiment, e.g., a human demonstrating actions for a robot. In this work we combine these approaches to teach a robot arm trajectories required for a number of interactive tasks.

Learning trajectory generating policies from demonstration has been addressed with for example a combination of Gaussian Mixture Models and Hidden Markov Models (Calinon et al., 2010), probabilistic flow tubes (Dong and Williams, 2011, 2012), or probabilistic motion primitives (Maeda et al., 2017b). The general strategy in this case is to first gather training data in the form of trajectories and to align these temporally e.g., with the help of Dynamic Time Warping (Sakoe and Chiba, 1978). Once the training data has been pre-processed in this way, the model of choice is trained to predict the trajectory of robotic motion for a certain task. During employment of the model, the online trajectory needs to be aligned with the temporal dynamics of the training samples in order to generate accurate movements. Depending on the trajectory representation, e.g., torque commands or Cartesian coordinates, the model's predictions might be highly dependent on the training data. For example, when the task is to learn to grasp an object at a certain location, the model might not generalize to grasping the same object at a different location.

This constant need of alignment and reliance on demonstrations hampers the models to work in a dynamic environment with varying task demands. For example, if the task is to shake hands with a human, the number of shaking cycles and the length of each individual shake can vary from trial to trial and have to be estimated online instead of being predicted once prior to the motion onset. These requirements for online interaction are discussed in more detail below.

2.2. Requirements for Online Interaction

Interaction with humans requires a robotic system to be flexible and adaptive (Dautenhahn, 2007; Maeda et al., 2017a). To meet these requirements, the robot needs to be able to anticipate future human actions and movements (Koppula and Saxena, 2015). Thus, imitation learning for interaction is different from non-social imitation learning as it requires to learn a function not only of one's own behavior, but also of the partner's behavior.

These requirements stand therefore in contrast to the approaches to imitation learning discussed in section 2.1 which focus on learning a trajectory of a fixed size. Maeda et al. (2017a) address the problem of adjusting to the speed of the human's actions by introducing an additional phase variable. This variable can be interpreted as an indication of the progress of the movement of the human to which the robot has to adapt.

FIGURE 1 | Kinesthetic teaching of a human-robot hand shake. The human partner is wearing a motion capture suit to record joint positions.

However, such an approach is only feasible for interactions which require little mutual adaptation beyond speed. For example, during a hand-shake interaction, it is not only important to meet the partner's hand at an appropriate time, but also to adjust to the frequency and height of every up-and-down movement. Thus, online interaction requires the prediction of the partner's future movements in order to adapt to them quickly and a continuous update of these predictions based on sensory feedback.

An additional requirement for natural human-robot interaction is to provide precise coordination. We envision a robot to actively engage in an interaction such that the human partner does not have to wait with a stretched arm until the robot reacts and moves its arm to engage in a hand-shake. Making use of predictive models allows the robot to initiate its movements before the human has reached the goal location. These models also provide a basis for collision-free path planning to assure safe interaction in shared workspaces.

Since humans are involved in the data collection process and kinesthetic teaching is time consuming and requires expert knowledge, the amount of training data is restricted. Therefore, any method used to learn trajectories must be data efficient. Many modern imitation learning techniques build on ideas from deep reinforcement learning (Li et al., 2017; Zhang T. et al., 2018) which is often data intensive. We rely on a model class which is regularized by its Bayesian foundation and therefore less prone to overfit to small datasets. This model class of deep latent variable models has been mostly used to model images. Here, we take inspiration from earlier work in which we model human motion trajectories (Bütepage et al., 2018a) and robot actions (Ghadirzadeh et al., 2017) with help of deep generative models. We extend the ideas to represent the dynamics of human-robot interaction in a joint model. For those unfamiliar with the ideas of Variational Autoencoders, we introduce the underlying concept of this model class below.

2.3. Deep Generative Models and Inference Networks

In this work, we model human and robotic motion trajectories with help of Variational Autoencoders (VAEs) (Rezende et al., 2014; Kingma and Welling, 2015), that is a class of deep generative models. In contrast to Generative Adversarial Networks (Goodfellow et al., 2014) and flow-based methods (Dinh et al., 2017; Kingma and Dhariwal, 2018), VAEs allow us to define our assumption in terms of a probabilistic, latent variable model in a principled manner. While we focus on the main concepts and the mathematical foundations of VAEs, we refer the reader to Zhang C. et al. (2018) for an in-depth review on modern advances in variational inference and VAEs. In the next section, we will shortly introduce the concepts of variational inference.

2.3.1. Variational Inference

To begin with, we assume that the observed variable, or data point, $\mathrm{x} \in \mathbb{R}^{d_x}$ depends on latent variables $\mathrm{z} \in \mathbb{R}^{d_z}$. If the dataset consists of images, the latent variables or factors z describe the objects, colors, and decomposition of the image. If, as we will introduce later, the dataset consists of human or robot joint movements, the underlying factors describe the general movement patterns. For example, a waving movement, in which many joints are involved, can be described by a single oscillatory latent variable. The dimension of z is smaller than the dimension of x, i.e., $d_z < d_x$, as it is a compressed representation of the observation. The precise size of the dimension is a modeling choice.

In general, this model describes a joint distribution over both variables $p_\theta(\mathrm{x}, \mathrm{z}) = p_\theta(\mathrm{x}|\mathrm{z})p_\theta(\mathrm{z})$ where θ are parameters. This modeling assumption allows us to generate new observations with help of the mathematical model instead of employing a physical system. First, a latent variable is sampled, from a prior distribution $\mathrm{z} \sim p_\theta(\mathrm{z})$. For example, to generate a waving arm movement, we sample where in the oscillation the arm starts and the initial velocity. Then we sample the actual poses conditioned on these latent variables. The conditional distribution $\mathrm{x} \sim p_\theta(\mathrm{x}|\mathrm{z})$ encodes the mapping from the latent space to the observed space. Thus, the generative process looks as follows:

$$\mathrm{x} \sim p_\theta(\mathrm{x}|\mathrm{z}),\ \mathrm{z} \sim p_\theta(\mathrm{z}). \tag{1}$$

In order to determine the structure of the latent variables that were generated on an observed set consisting of n data points $\mathrm{X} = \{\mathrm{x}_i\}_{i=1:n}$, one requires access to the posterior distribution $p_\theta(\mathrm{z}_i|\mathrm{x}_i)$ for each data point x_i. Often exact inference of this term is intractable which is why one recedes to approximate inference techniques such as Monte Carlo sampling and variational inference (VI). VAEs combine VI for probabilistic models with the representational power of deep neural networks. VI is an optimization based inference technique which estimates the true posterior distribution $p_\theta(\mathrm{Z}|\mathrm{X})$ with a simpler approximate distribution $q_\phi(\mathrm{Z})$ where ϕ are parameters and $\mathrm{Z} = \{\mathrm{z}_i\}_{i=1:n}$ is the set of latent variables corresponding to the data set. A common approach is the mean-field approximation which assumes that the latent variables are independent of each other

$q_\phi(\mathrm{Z}) = \prod_{i=1}^{n} q_\phi(\mathrm{z}_i)$. As an example, if $q_\phi(\mathrm{z}_i)$ follows a Gaussian distribution, we need to identify a mean μ_i and variance σ_i for every data point $q_\phi(\mathrm{Z}) = \prod_{i=1}^{n} \mathcal{N}(\mu_i, \sigma_i)$. For the entire dataset (X, Z), the parameters of this distribution are determined by optimizing the Evidence Lower BOund (ELBO).

$$\log p_\theta(\mathrm{X}) \geq \mathrm{E}_{q_\phi(\mathrm{Z})} \log \frac{p_\theta(\mathrm{X}, \mathrm{Z})}{q_\phi(\mathrm{Z})}$$
$$= \mathrm{E}_{q_\phi(\mathrm{Z})} \log p_\theta(\mathrm{X}|\mathrm{Z}) - D_{KL}(q_\phi(\mathrm{Z})||p_\theta(\mathrm{Z})), \qquad (2)$$

where the Kullback–Leibler divergence $D_{KL}(q||p) = \mathrm{E}_q \log \frac{q}{p}$ is a distance measure between two distributions q and p.

Traditional VI approximates a latent variable distribution $q_\phi(\mathrm{z}_i)$ for every data point i which becomes expensive or impossible when the number of data points n is large. VAEs circumvent this problem by learning a direct functional mapping from the data space to the latent space and vice versa. We will detail this method in the next section.

2.3.2. Variational Autoencoders

Instead of approximating n sets of parameters, VAEs employ so called inference networks to learn a functional mapping from the data space into the latent space. Thus, we define each latent variable to be determined by a distribution $\mathrm{z}_i \sim q_\phi(\mathrm{z}_i|\mathrm{x}_i)$ which is parameterized by a neural network (the inference network) that is a function of the data point x_i. In the Gaussian case this would imply that $\mathrm{z}_i \sim \mathcal{N}(\mu(\mathrm{x}_i), \sigma(\mathrm{x}_i))$, where $\mu(\cdot)$ and $\sigma(\cdot)$ are neural networks mapping from the data space to the parameter space of the latent variables. Likewise, the likelihood is represented by neural network mappings (the generative network) $\mathrm{x}_i \sim p_\theta(\mathrm{x}_i|\mathrm{z}_i)$. In this way, VAEs do not estimate n sets of parameters but only the parameters of the inference and generative network. These are optimized with help of the ELBO

$$\log p_\theta(\mathrm{X}) \geq \mathcal{L}(\mathrm{X}, \theta, \phi)$$
$$= \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}_{q_\phi(\mathrm{z}_i|\mathrm{x}_i)} \log p_\theta(\mathrm{x}_i|\mathrm{z}_i) - D_{KL}(q_\phi(\mathrm{z}_i|\mathrm{x}_i)||p_\theta(\mathrm{z}_i)). \qquad (3)$$

Note that we replaced the expectation in Equation (2) with the Monte Carlo estimate summing over the individual data points.

3. METHODOLOGY

Following the introduction of VAEs above, we will now detail how we employ them to learn the sensorimotor patterns required for interactive tasks. We will begin with a description of human-human dynamics modeling which is subsequently used to guide the human-robot interaction model.

3.1. A Generative Model of Interaction

In general we assume that a recording rec consists of T_{rec} observations $\mathrm{x}^{s_1}_{1:T_{rec}}$ and $\mathrm{x}^{s_2}_{1:T_{rec}}$, where $(s_1, s_2) = (human_1, human_2)$, and x^s_t represents a single frame containing the joint positions of human $s \in \{s_1, s_2\}$. During testing time, we would like to be able to infer a future window (of size w) of observations after time t, i.e., we would like to predict $\mathrm{x}^{s_1}_{t:t+w}$ and $\mathrm{x}^{s_2}_{t:t+w}$. We assume a generative process that looks as follows

$$\mathrm{x}^{s_1}_{t:t+w} \sim p_{\theta_x}(\mathrm{x}^{s_1}_{t:t+w}|\mathrm{z}^{s_1}_t), \quad \mathrm{z}^{s_1}_t \sim p_{\theta_z}(\mathrm{z}^{s_1}_t|\mathrm{d}_t),$$
$$\mathrm{d}_t \sim p_{\theta_s}(\mathrm{d}_t|h^{s_1}_t), \quad h^{s_1}_t = f_\psi(h^{s_1}_{t-1}, \mathrm{x}^{s_1}_{t-1}) \qquad (4)$$
$$\mathrm{x}^{s_2}_{t:t+w} \sim p_{\theta_x}(\mathrm{x}^{s_2}_{t:t+w}|\mathrm{z}^{s_2}_t), \quad \mathrm{z}^{s_2}_t \sim p_{\theta_z}(\mathrm{z}^{s_2}_t|\mathrm{d}_t),$$
$$\mathrm{d}_t \sim p_{\theta_s}(\mathrm{d}_t|h^{s_2}_t), \quad h^{s_2}_t = f_\psi(h^{s_2}_{t-1}, \mathrm{x}^{s_2}_{t-1}). \qquad (5)$$

Here, the latent variables $\mathrm{z}^{s_1}_t$ and $\mathrm{z}^{s_2}_t$ for agent s_1 and s_2 encode the next time window $\mathrm{x}^{s_1}_{t:t+w}$ and $\mathrm{x}^{s_2}_{t:t+w}$, while $h^{s_2}_t$ is the deterministic output of a recurrent model f_ψ. The role of $h^{s_2}_t$ is to summarize the information contained in the past observations $t' < t$, which in turn is transformed into the shared task dynamics d_t. From a system perspective, d_t is the stochastic output of a neural network that driven by the hidden state vector $h^{s_2}_t$. As depicted in **Figure 2**, the d_t can be derived from the movement of either subject independently. These shared dynamics influence how each partner moves through $\mathrm{z}^{s_1}_t$ and $\mathrm{z}^{s_2}_t$. In summary, the generative model for agent s_1 represents the joint distribution $p_\theta(\mathrm{x}^{s_1}_{t:t+w}, \mathrm{z}^{s_1}_t, \mathrm{d}_t|h^{s_1}_t)$ conditioned on a deterministic summary of the past $h^{s_1}_t$ and parameterized by $\theta = (\theta_x, \theta_z, \theta_s)$.

In the following, we will describe how to learn each of the components for human-human interaction and subsequently how to transfer this knowledge to a human-robot interaction scenario.

3.1.1. Motion Embeddings

One problem when it comes to predicting the future is that there exist many possible ones. When using a mean-squared error based cost function during training, this will lead the model to rely on predicting only the average, not many different trajectories. We will circumvent this problem by first learning a latent space that encodes the future without knowledge of the past and then to learn a distribution over the latent variables which is conditioned on the past [e.g., $p(\mathrm{z}^{s_1}_t|\mathrm{d}_t)$ in Equation 5]. At each time step, we assume that there exists latent variables $\mathrm{z}^{s_1}_t$ and $\mathrm{z}^{s_2}_t$ for agent s_1 and s_2 which encode the next time window $\mathrm{x}^{s_1}_{t:t+w}$ and $\mathrm{x}^{s_2}_{t:t+w}$. We assume that both humans are encoded into a common space, therefore we will replace the superscripts s_1 and s_2 with s in the following discussion.

To infer the latent variables efficiently from data, we apply variational autoencoders (introduced in section 2.3). To this end, we define the following generative process:

$$\mathrm{x}^s_{t:t+w} \sim p_{\theta_x}(\mathrm{x}^s_{t:t+w}|\mathrm{z}^s_t),$$
$$\mathrm{z}^s_t \sim p_{\theta_z}(\mathrm{z}^s_t)$$
$$= \mathcal{N}(0, 1), \text{ and approximate posterior } \mathrm{z}^s_t \sim q_{\phi_z}(\mathrm{z}^s_t|\mathrm{x}^s_{t:t+w}). \qquad (6)$$

The graphical model is depicted in **Figure 3A**. The parameters (θ_x, ϕ_z) of the generative network $p_{\theta_x}(\mathrm{x}^s_{t:t+w}|\mathrm{z}^s_t)$ and the inference network $q_{\phi_z}(\mathrm{z}^s_t|\mathrm{x}^s_{t:t+w})$ are jointly trained on the training data collected from both humans to optimize the Evidence Lower BOund (ELBO).

$$\mathcal{L}(\mathrm{x}^s_{t:t+w}, \theta, \phi) = \mathbb{E}_{q_{\phi_z}(\mathrm{z}^s_t|\mathrm{x}^s_{t:t+w})} \log p_{\theta_x}(\mathrm{x}^s_{t:t+w}|\mathrm{z}^s_t)$$
$$-D_{KL}(q_{\phi_z}(\mathrm{z}^s_t|\mathrm{x}^s_{t:t+w})||p_{\theta_x}(\mathrm{z}^s_t)). \qquad (7)$$

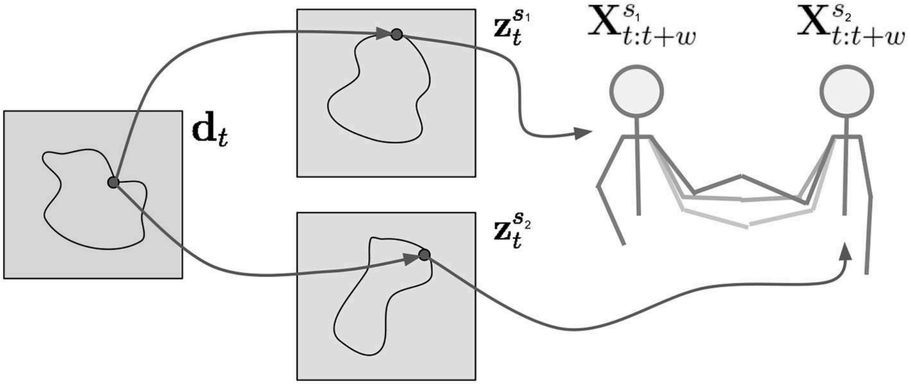


FIGURE 2 | The task dynamics d_t govern the activity of the latent variables of both partners $z_t^{s_1}$ and $z_t^{s_2}$. These in turn determine the future movement of the partners $x^{s_1}_{t:t+w}$ and $x^{s_2}_{t:t+w}$.

3.1.2. Encoding Task Dynamics

Once the motion embeddings have been learned, the whole generative model in Equation (5), as depicted in **Figure 3B**, can be trained. To this end, we need to infer the parameters $(\theta_z, \theta_s, \psi)$ to estimate $p_{\theta_z}(z^s_t|d_t), p_{\theta_s}(d_t|h^s_t)$ and $f_\psi(h^s_{t-1}, x^s_{t-1})$.

The loss function is defined as follows

$$\begin{aligned}&\mathcal{S}(x^{s_1}_{t-1:t+w}, x^{s_2}_{t-1:t+w}, \theta_z, \theta_s, \psi)\\&= \sum_{s\in\{s_1,s_2\}} D_{KL}(p_{\theta_z}(z^s_t|d_t)||q_{\phi_z}(z^s_t|x^s_{t:t+w})) +\\&JSD(p_{\theta_s}(d_t|h^{s_1}_t)||p_{\theta_s}(d_t|h^{s_2}_t)).\end{aligned} \tag{8}$$

The first line in Equation (8) forces the distributions over latent variables z^s_t that depend on the past to be close to the expected motion embedding at time t. The second line enforces that the latent variable d_t, which encodes the task dynamics are the same for both agents. As the KL divergence is not symmetric, we use here the Jensen–Shannon divergence, which is defined as $JSD(p||q) = \frac{1}{2}(D_{KL}(p||\frac{1}{2}(p+q)) + D_{KL}(q||\frac{1}{2}(p+q)))$ for two distributions p and q.

3.1.3. Interactive Embodiment Mapping

Once trained, the generative model described above can be used to generate future trajectories for both agents given that only one agent has been observed. This is achieved by e.g., predicting the task dynamics variable $d_t \sim p_{\theta_s}(d_t|h^{s_1}_t)$ with help of data collected for agent s_1 and using this variable to infer both $z^{s_1}_t \sim p_{\theta_z}(z^{s_1}_t|d_t)$ and $z^{s_2}_t \sim p_{\theta_z}(z^{s_2}_t|d_t)$. We will make use of this fact to infer not only a human partner's future movement, but also to guide how a robotic partner should react given the observed human.

As training data acquisition with a robot and a human in the loop is cumbersome and time consuming, we do not have access to as much training data of the human-robot interaction compared to the human-human interaction. Therefore, we will leverage the task dynamics representation learned from human-human interaction to guide the robot's corresponding motion commands. To this end, we extract the task dynamics distribution from the human partner for each time step of the human-robot interaction recordings and learn a mapping to the robot's motion commands with a second dynamics model.

In more detail, given a recording *rec* which consists of T_{rec} observations $x^{s_1}_{1:T_{rec}}$ and $x^r_{1:T_{rec}}$, where x^r_t represents the robot's state at time t, we first collect $d_{1:T_{rec}}$ which we set to the mean of the distribution $p_{\theta_s}(d_t|h^{s_1}_t)$ for each time step t. We are now equipped with a training data set, containing the data point pairs $(x^r_{t:t+w}, d_t)$. In order to learn a predictive model from the task dynamic variable d_t to the future motion commands of the robot, $x^r_{t:t+w}$, we design a similar approach to the model described for human-human interaction. It includes a Variational Autoencoder functioning as a motion embedding and a recurrent network that encodes the robot motion over time. These two models are depicted in **Figures 3C,D**, respectively.

3.1.3.1. Interaction model with predictive input

Similar to the human-human setting in Equation (5), the generative model for the robot motion is as follows

$$\begin{aligned}x^r_{t:t+w} &\sim p_{\theta_{xr}}(x^r_{t:t+w}|z^r_t), \quad z^r_t \sim p_{\theta_{zr}}(z^r_t|h^r_t),\\ h^r_t &= f_{\psi_r}(h^r_{t-1}, x^r_{t-1}, d_{t-1}).\end{aligned} \tag{9}$$

Just as in the human-human setting, we first train a motion embedding VAE on the robot data, i.e., we train the following model with the same loss function as in Equation (7)

$$\begin{aligned}&x^r_{t:t+w} \sim p_{\theta_{xr}}(x^r_{t:t+w}|z^r_t), \quad z^r_t \sim p_{\theta_{zr}}(z^r_t)\\&= \mathcal{N}(0,1), \text{ and approx. posterior } z^r_t \sim q_{\phi_{zr}}(z^r_t|x^r_{t:t+w}).\end{aligned} \tag{10}$$

Subsequently, we assume that the parameters (θ_{zr}, ψ_r) in Equation (9) are inferred by optimizing

$$\mathcal{S}(x^{r_1}_{t-1:t+w}, d_t, \theta_z, \psi) = D_{KL}(p_{\theta_{zr}}(z^s_t|h^r_t)||q_{\phi_{zr}}(z^s_t|x^r_{t:t+w})), \tag{11}$$

where the dynamics $d_t \sim p_{\theta_s}(d_t|h^{s_1}_t)$ are extracted with help of the models trained on the human-human data. We summarize the training procedure of all our model in Algorithm 1.

Algorithm 1: All four steps of our combined motion embedding and dynamics modeling framework.

```
Human-human interaction
    Data: $x^{s_1,s_2} = \{x^{s_1}_{1:T_{rec}}, x^{s_2}_{1:T_{rec}}\}_{rec \in \text{HHI recordings}}$
    Step 1: Human motion embedding
        Fit $p_{\theta_x}(x^s_{t:t+w}|z^s_t)$ and $q_{\phi_z}(z^s_t|x^s_{t:t+w})$ to $x^{s_1,s_2}$,
        following Equation 7.
    Step 2: Task dynamics
        Fit $p_{\theta_z}(z^s_t|\mathrm{d}_t), p_{\theta_s}(\mathrm{d}_t|h^s_t)$ and $f_\psi(h^s_{t-1}, x^s_{t-1})$ to $x^{s_1,s_2}$,
        following Equation 8.
Human-robot interaction
    Data: $x^{s_1,r} = \{x^{s_1}_{1:T_{rec}}, \mathrm{d}_{1:T_{rec}}, x^r_{1:T_{rec}}\}_{rec \in \text{HRI recordings}}$,
        where $\mathrm{d}_t =$ mean of $p_{\theta_s}(\mathrm{d}_t|h^s_t)$
    Step 3: Robot motion embedding
        Fit $p_{\theta_{xr}}(x^r_{t:t+w}|z^r_t)$ and $q_{\phi_{zr}}(z^r_t|x^r_{t:t+w})$ to $x^{s_1,r}$,
        combining Equation 7 and 10.
    Step 4: Interactive embodiment mapping
        Fit $p_{\theta_{zr}}(z^r_t|h^r_t)$ and $f_{\psi_r}(h^r_{t-1}, x^r_{t-1}, \mathrm{d}_{t-1})$ to $x^{s_1,r}$,
        following Equation 11.
```

3.2. Generating Interactions

In order to employ our models during an ongoing interaction, we need to predict future time steps. As the dynamics and the motion embeddings encode a window of the next w time steps, the prediction up to this horizon is straight forward as it only requires a propagation of the observed data. To go beyond a time frame of w is made possible by our generative design. Instead of propagating observed data, one can let the models predict the next w time frames based on the observed data and provide these as an input to the model. In case of the human-robot interaction model, one has to first predict the human's future motion to extract the matching dynamics variables and can subsequently use these variables together with predictions of the robot's motion to generate long-term robot motion. During online interaction these predictions can be updated on the fly when new data is observed.

3.3. Baselines

We benchmark our approach on three baselines. Our own approach will be called *Human Motion Embedding* in the following.

The first baseline tests whether our predictive and adaptive approach is necessary or whether more static imitation learning techniques suffice. To test this, we group the robot trajectories in the training data according to action type and use Dynamic Time Warping (DTW) to align them. We fit Gaussian distributions with full covariance matrices to the trajectory of each of the robot's joints. If DTW resulted in a trajectory length of T_{DTW} for a certain action type and joint, then the Gaussian is of dimension T_{DTW}. A sample from each Gaussian model constitutes therefore a trajectory in joint angle space without input from the current human movement. We call this approach *Gaussian model*.

The second baseline tests whether our approach actually benefits from the encoded dynamics learned with the HHI data. Thus, in this case we train the same model as described in section 3.1.3. However, instead of feeding the dynamics variable d_t into the recurrent network $h^r_t = f_{\psi_r}(h^r_{t-1}, \mathrm{x}^r_{t-1}, \mathrm{d}_{t-1})$ in Equation (9), we feed the current human joint position x^s_{t-1}, i.e., $h^r_t = f_{\psi_r}(h^r_{t-1}, \mathrm{x}^r_{t-1}, \mathrm{x}^s_{t-1})$. This also affects the loss in Equation (11), which now is a function of x^s_{t-1}, i.e., $\mathcal{S}(\mathrm{x}^{r_1}_{t-1:t+w}, \mathrm{x}^s_{t-1}, \theta_{zr}, \psi_r)$. We call this approach *Raw Data HR* which symbolizes that we provide raw human and robot data as input to the model.

The third baseline tests whether the human observation is required at all or whether the approach is powerful enough to predict based on robot joint position alone. In this case we train the same model as described in section 3.1.3, but provide only the current robot joint positions x^r_{t-1}, i.e., $h^r_t = f_{\psi_r}(h^r_{t-1}, \mathrm{x}^r_{t-1})$. This also affects the loss in Equation (11), i.e., $\mathcal{S}(\mathrm{x}^{r_1}_{t-1:t+w}, \theta_{zr}, \psi_r)$. We call this approach *Raw Data R* which symbolizes that we provide only raw robot data as input to the model.

4. EXPERIMENTAL SETUP AND MODELS

In this section we describe the experimental setup as well as modeling decisions and the model training procedure. For more details regarding model architectures and model training, such as train and test splits (please see the **Supplementary Material**).

4.1. Task Description

Our interactive tasks consist of performing four different greeting gestures with a human partner. In each task execution we assume the identity of the gesture to be known apriori as the focus of this work lies on learning continuous interactive trajectories. However, our method can easily be extended to automatically infer the action type (Bütepage et al., 2018b). Two of the gestures fall into the category of dyadic leader–follower interaction, while the other two partners carry equal roles. The interactive gestures are detailed in **Table 1**. Between actions, the two partners are standing in an upright position with both arms directed downwards close to the body.

As the robot is not necessarily equipped with a hand-like gripper, the actions involving finger movement are omitted during human-robot interaction. Furthermore, we assume the robot to take the role of the follower.

4.2. Data Collection

We collected data from human-human and human-robot interaction, respectively. The robotic setup and the human motion recording setup are described below, followed by the data collection procedure.

4.2.1. Robotic System Setup

In this work, we use a dual-armed YuMi-IRB 14000 robot which has been developed by ABB specifically with human-robot collaboration in mind. As depicted in **Figure 4A**, each arm has seven joints Arm 1 (rotation motion), Arm 2 (bend motion), Arm 7 (rotation motion), Arm 3 (bend motion), Wrist 4 (rotation motion), Wrist 5 (bend motion), and Flange 6 (rotation motion). To control the robot, we work in the joint angle space, i.e., at each time step we have access to a seven dimensional vector consisting of radial measurements. To control the robot, we provide the system with the next expected joint angle configuration or a

TABLE 1 | Gesture descriptions for both equal and leader-follower roles.

Equal roles	
Hand waving:	Both: Lifting the right arm into an upright, 90-degree angle with the open palm facing the partner; moving the lower arm sideways in an oscillatory motion (around 3–6 cycles); lowering the arm.
Hand Shaking:	Both: Stretching the right arm forward to meet the partner's hand, grasping the partners hand; moving the lower arm up and down in an oscillatory motion (around 3–6 cycles); releasing the partner's hand, lowering the arm.
Leader-follower roles	
Parachute fist-bump:	Both: Stretching the right arm upwards with the hand closed to a fist to meet the partner's hand, touching the partner's fist with one's own; Leader: (parachute) Opening the hand and tilting it so that the flat, inner palm faces downwards; keeping the hand above the follower's hand; moving the hand in a slight sideways oscillatory motion while simultaneously moving downwards; Follower (person): Keeping the hand closed and slightly below the leader's hand; following the slight sideways oscillatory motion of the leader and moving the hand downwards; Both: Lowering the arm when the hand is approximately on the height of the hip.
Rocket fist-bump:	Both: Stretching the right arm downwards with the hand closed to a fist to meet the partner's hand, touching the partner's fist with one's own; Leader (rocket): Opening the hand slightly to point to fingers upwards; keeping the hand above the follower's hand; moving the hand upwards; Follower (fire): Opening the hand with all fingers stretched downwards; keeping the hand below the leader's hand; wiggling the fingers to simulate fire; moving the hand upwards; Both: Lowering the arm when the hand is approximately on the height of the shoulders.

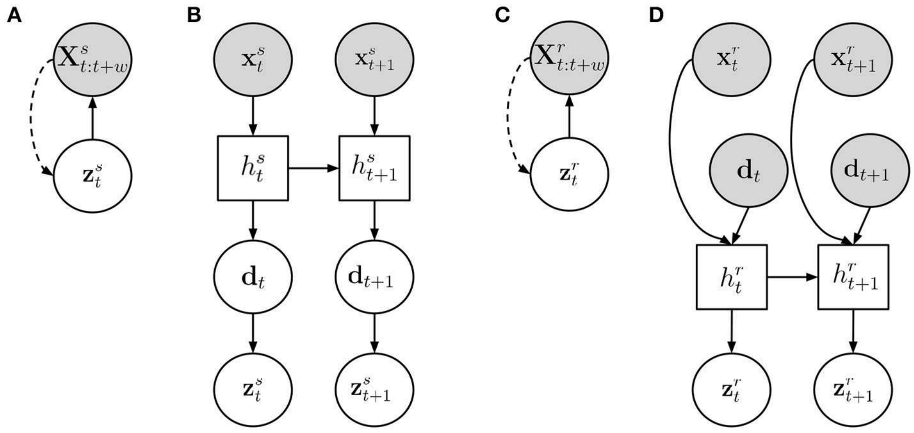


FIGURE 3 | An overview of the model structure. **(A)** Human motion embedding, **(B)** Task dynamics model, **(C)** Robot motion embedding, **(D)** Human motion embedding, Interaction model with predictive input. Gray circles represent observed variables, white circles are unobserved variables, and a white square indicates a deterministic unit. A filled line shows the generative process while dotted lines indicate inference connections.

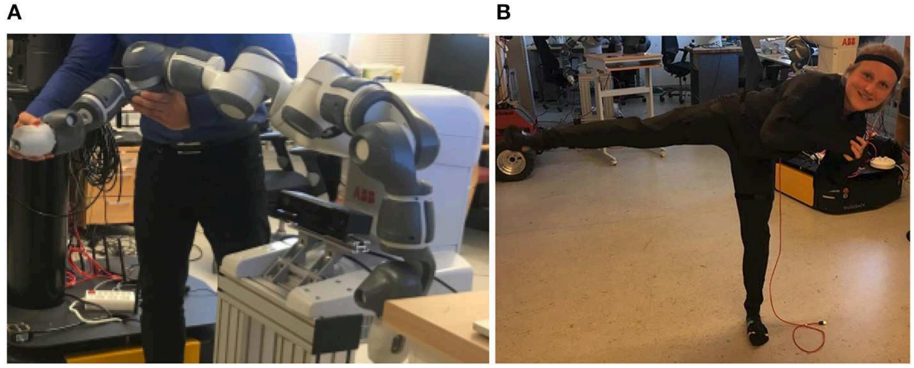


FIGURE 4 | (A) The right arm of the Yumi robot used in the experiments. **(B)** A rokoko smart suit in action.

TABLE 2 | Statistics of the collected dataset.

	Human-human data			Human-robot data		
Action type	**# Trials**	**Min. duration (s)**	**Max. duration (s)**	**# Trials**	**Min. duration (s)**	**Max. duration (s)**
Hand shake	38	8.5	12.5	10	10.4	14.5
Hand wave	31	8.5	17.5	10	12.7	17.4
Parachute	49	7.0	12.0	11	11.0	14.3
Rocket	70	3.0	6.0	10	11.1	13.8

whole trajectory thereof. We sample the robot's joint angles at a frequency of 40 Hz.

4.2.2. Human Motion Capture

We recorded the 3D position of the human joints in Cartesian space during interaction with help of two Rokoko smart suits[*]. As shown in **Figure 4B**, these textile suits are equipped with 19 inertia sensors with which motion is recorded. Via wireless communication with a Wi-fi access point, the suits are able to record whole-body movements at a rate of up to 100 Hz. While simultaneous recordings with several suits are possible, we align the recordings offline. We record the 3D Cartesian positions of each joint in meters with respect to a body-centric reference frame. The data is sampled down to match the 40 Hz of the robot recording.

4.2.3. Collection Procedure

For the human-human dataset, we asked two participants to perform all four actions as described in section 4.1 for approximately 6 min each. The exact number of repetitions of each action type as well as duration statistics are listed in **Table 2**. A recording of the action *hand-shake* is depicted in the top of **Figure 5**.

For the robot-human dataset, we asked one of the participants to perform all four actions together with the robot. To this end, we made use of kinesthetic teaching, i.e., a human expert guided the arm of the robot during the interaction. As shown in **Table 2**, the duration of the human-robot trials is on average slightly longer than the human-human trials. A recording of the action *hand-shake* is depicted in the middle of **Figure 5**.

4.3. Modeling Decisions and Training Procedure

All models are implemented in Tensorflow (Abadi et al., 2015). Instead of training four separate models, one for each action, we train a single model that can generate all actions. In order to signal to the model, which action is currently performed, we encode the actions as a one-hot vector which is passed as an additional input to the model as described below.

4.3.1. Modeling Choices

All latent variables $(z_t^{s_1}, z_t^{s_2}, z_t^r, d_t^s)$ are chosen to be independent and identically distributed Gaussian units with a trainable mean and variance. The prior of the VAEs is set to be standard normal distributed $p_{\theta_z}(z_t) \sim \mathcal{N}(0, 1)$.

To indicate to the recurrent models which action is currently performed, we provide the networks with a one-hot vector indicating the current action. We add an additional *not-active* action, which indicates those time steps after completion of the interaction. Thus, the one-hot vector is of dimension 5 and is concatenated with the observed joint positions of either human or robot.

We train two identical models for the two human partners while the model of the robot motion has a different structure. Please see the **Supplementary Material** for details about model architecture.

4.3.2. Data Representation

We represent the human by four joints "RightShoulder," "RightArm," "RightForeArm," and "RightHand" in 3D Cartesian space, resulting a 12 dimensional vector. We center the arm around the shoulder joint. The robot is represented by a seven dimensional vector, each indicating a joint angle. We select 80% of all trials of a certain interaction as training data and keep 20% as testing data. In practice, we keep the last 20% of trials of the recording. This results in 149 trials as training data and 39 trials of testing data for the HHI recordings in 32 trials as training data and 9 trials of testing data for the HRI recordings.

4.3.3. Training Details

For optimization we use the Adagrad optimizer with a learning rate of 0.001. The batch size is 5,000 for the VAEs and 500 for the recurrent networks. If a dataset does not contain that many samples, we replicate the training samples to get to 5,000. We train all models until convergence. For the VAEs we use a form of β-VAE (Higgins et al., 2017), where $\beta = 0.5$. For training the recurrent networks, we pad all data sequences with ones to have the same length.

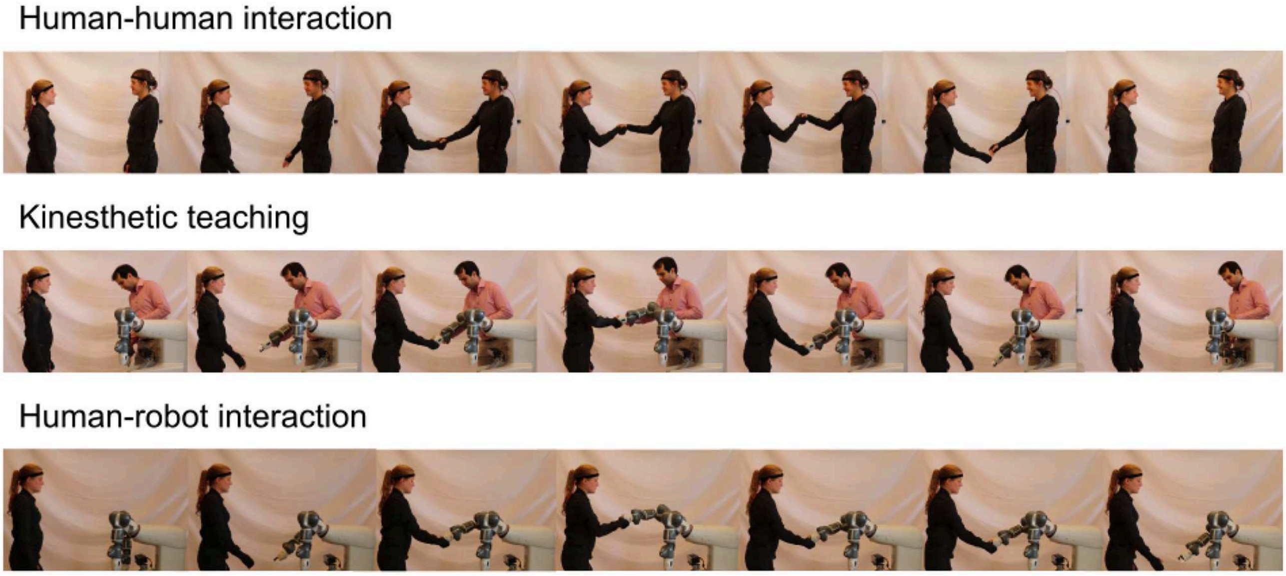
5. RESULTS

In this section we present the performance of the proposed approach. Online employment of our approach during the action *hand-shake* is depicted in the bottom of **Figure 5**. More examples can be found in the Supplementary Material in form of a video (**Supplementary Video 1**). In the analysis we present results on held-out test datasets. As described in section 4.3.1, each model was trained on all actions simultaneously and subsequently tested on each of the actions in the held-out test dataset.

We begin by investigating the predictive performance of the models trained on the human-human dataset. This will be followed by an analysis of the robot motion prediction. In this case, we consider both the predictive error as well as the entrainment of predicted vs. ground truth robot motion with the human motion.

---

[*] https://www.rokoko.com/



Human-human interaction

Kinesthetic teaching

Human-robot interaction

FIGURE 5 | The data collected during human-human interaction **(top)** and kinesthetic teaching **(middle)** is used train the proposed models. These are employed in human-robot interaction tasks **(bottom)**.

5.1. Predictive Performance on Human-Human Data

We have two reasons for collecting additional human-human interaction data. Firstly, we hypothesize that the dynamics learned based on HHI data can guide robot action selection during HRI experiments. Secondly, it is easier to collect HHI data, allowing for larger datasets. To test the second hypothesis we trained the human motion embedding and dynamics models both on HHI data and only on the human data contained in the HRI data. In the latter case, the dynamics variable is not restricted to match a human partner. We test the predictive capacity of both these models by computing the mean squared prediction error (MSPE) for the time window *w* on both test data sets (HRI and HHI). The results are depicted in **Figure 6**. Two observations can be made. First of all, the model trained on HRI data does not generalize well, mainly caused by the small training data set. Secondly, the prediction error does not increase drastically over time as should be expected. Due to the fact that we do not force the model to predict a whole trajectory as e.g., (Bütepage et al., 2018a) but only a latent variable which can be decoded into a trajectory, our model is less prone to regress to the mean but to encode the actual motion.

TABLE 3 | NRMSD computed on robot testing data averaged over all joints.

Human motion embedding	Raw data RH	Raw data R	Gaussian model
NRMSD computed on robot testing data			
0.16	0.22	0.18	0.20

The bold value indicates which model performed best.

5.2. Predictive Performance on Human-Robot Data

In this section we inspect how our proposed dynamics transfer approach performs against the baselines. As the different joints move to different extents, the range of joint angles varies. Therefore, we measure the predictive error not with the MSPE as in the case of HHI predictions but with the normalized root-mean-square deviation (NRMSD) which is computed as follows:

$$NRMSD(\{x^r_{1:T_{tr},j}\}_{tr\in 1:TR}, \{\hat{x}^r_{1:T_{tr},j}\}_{tr\in 1:TR}) = \frac{1}{TR}\sum_{tr\in 1:TR}\sqrt{\frac{1}{T_{tr}(j_{max}-j_{min})}\sum_{t=1}^{T_{tr}}(x^r_{t,j}-\hat{x}^r_{t,j})^2}, \quad (12)$$

for the *jth* joint. Here *tr* denotes trial, i.e., one execution of an interaction, and *TR* is the number of trials. j_{max} and j_{min} denote the maximum and minimum value that has been recorded for the *jth* joint in the training data.

We start by comparing our approach (Human Motion Embedding) to the two models that have an identical structure but that differ in the type of input data (Raw Data HR and Raw Data R). To this end, we provide ten time steps as input to the models and let the recurrent network predict 30 steps as described in section 3.2. This process is repeated until the end of a trial is reached. Since the Raw Data HR model is not able to generate human motion, we provide it with the last observed human pose. Through the motion embedding, the models produce a prediction of the next 40 time steps (1 s). We average over all time steps and present the results in **Figure 7**. The Human Motion Embedding appears to produce the smallest errors, especially for those joints that are vital for the interaction

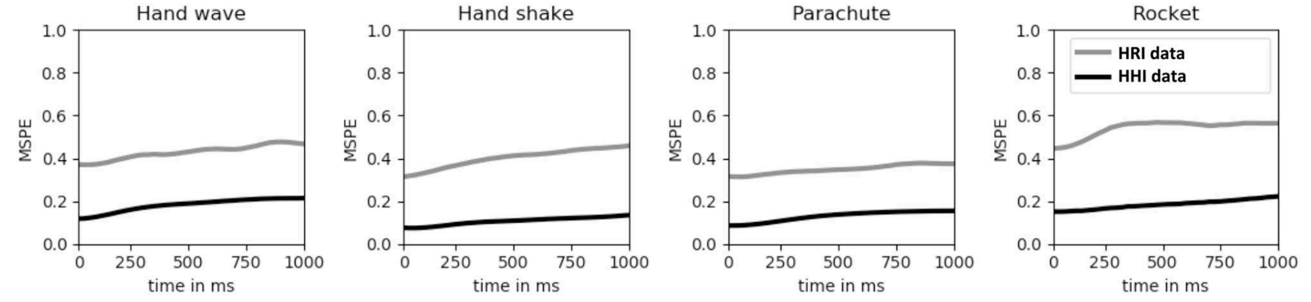


FIGURE 6 | The mean squared prediction error (MSPE) in meters for human-human interaction over a time horizon of 1 s. The human motion embedding and dynamics models were trained on the HHI data and on the human data contained in the HRI data set. In the latter case, the dynamics variable is not restricted to match a human partner.

(joint 2, 3, and 4). The wrist joints (joint 6 and 7) are of less importance and do also show a larger degree of between-trial variance in the training data. We depict the predictions of each of the Human Motion Embedding model, the Raw Data HR model and the ground truth trajectory for one testing trial of each action in **Figure 8**.

When averaged over the forty time steps of prediction, the difference becomes clear in **Table 3**, where we also include the Gaussian model. As the Raw Data HR model is not able to predict human motion, it produced the largest error. The Human Motion Embedding outperforms both the adaptive Raw Data HR and Raw Data R models as well as the non-adaptive Gaussian model. The adaptive Raw Data R model produces a smaller error than the non-adaptive Gaussian model, which also is trained on raw robot data. We will investigate the difference between adaptive and non-adaptive approaches in more detail in the next section.

5.3. Non-adaptive vs. Adaptive Motion Generation

As discussed in section 2.2, Human-Robot interaction has additional requirements compared to traditional imitation learning. It does not suffice to learn a distribution over the trajectories observed in the training data and sample a whole trajectory during run-time. Instead, HRI requires adaptive and predictive models that react to the human's actions such that a sensorimotor coupling between human and robot can arise. We visualize this in **Figure 9** by sampling from the Gaussian model of joint 4 for the action hand-shake. It becomes apparent that none of the samples is in accordance with any of the testing trials that are also depicted. First of all, the motion onset differs and the duration of the trajectory is predetermined due to the time alignment, while the duration of natural interaction differs from trial to trial. Additionally, the movement is not adapted to the human's hand-shake but has different degrees of phase shift. If we compare these predictions to the predictions of joint 4 in the second row of **Figure 8**, we realize that the adaptive approach reacts in a timely manner and follows the oscillations of the ground truth motion that match the human motion. We will investigate the degree of entrainment of the predictions of robot with the human motion in the next section.

5.4. Entrainment on Human-Robot Data

With this work we are aiming at developing models that allow for sensorimotor coupling between humans and robots to benefit physical HRI. We visualize the generated predictions of the Human Motion Embedding model as well as the ground truth robot motion data and the hand position of the human for a testing trial of each interaction in **Figure 10**. As not all joints are relevant to a task, we visualize joint 2–4. We see that the predicted motion follows the oscillatory movement of the human hand during hand-wave (see joint 3), hand-shake (see joint 4), and the parachute (see joint 4) interaction as well the rise and fall of the rocket action (see joint 2 and 3).

To investigate whether the models capture this coupling, we extract the dynamics variables of the human motion of an entire testing trial of the hand-shake interaction as well as the latent variables that predict the robot motion. We then apply factor analysis to these two streams of data and compare the two first components to each other. The two components are visualized in **Figure 11**. The first factor appears to represent the general onset, duration and offset of the interaction while the second factor encodes the oscillatory motion of the hand and arm. We see that, although the factor analysis is performed independently on the human and robot latent variables, the overall structure is similar. Additionally, the oscillatory motion is overlapping, indicating a coupling between the two systems.

6. CONCLUSION

In this work, we propose a deep generative model approach to imitation learning of interactive tasks. Our contribution is a novel probabilistic latent variable model which does not predict in joint space but in latent space, which minimizes the chance of regression to the mean. We employ this model both as a dynamics extractor of HHI as well as the basis for the motion generation of a robotic partner. Our experiments indicate that HRI requires adaptive models which take the human motion and task dynamics into account. These dynamics, which encode the movement of both humans (see **Figure 2**), and therefore the

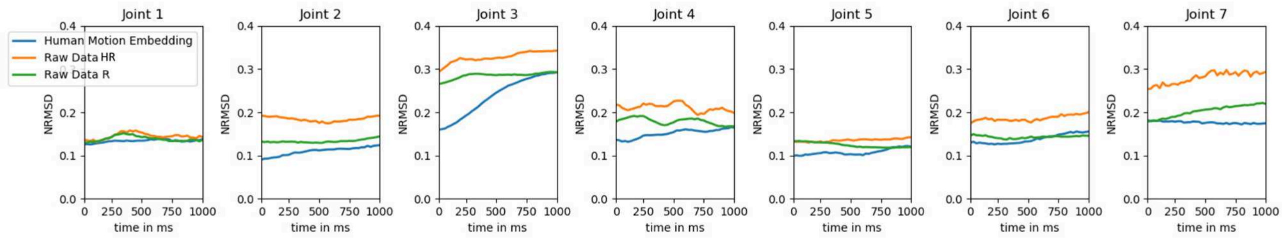


FIGURE 7 | The normalized root-mean-square deviation (NRMSD) for robot motion during human-robot interaction over a time horizon of 1 s. The error is averaged over the time steps of all trials. The models are provided with 10 time steps as input and the recurrent network predict 30 steps as described in section 3.2. This process is repeated until the end of a trial is reached. The Raw Data HR model receives the last observed human pose as it is not able to generate human motion.

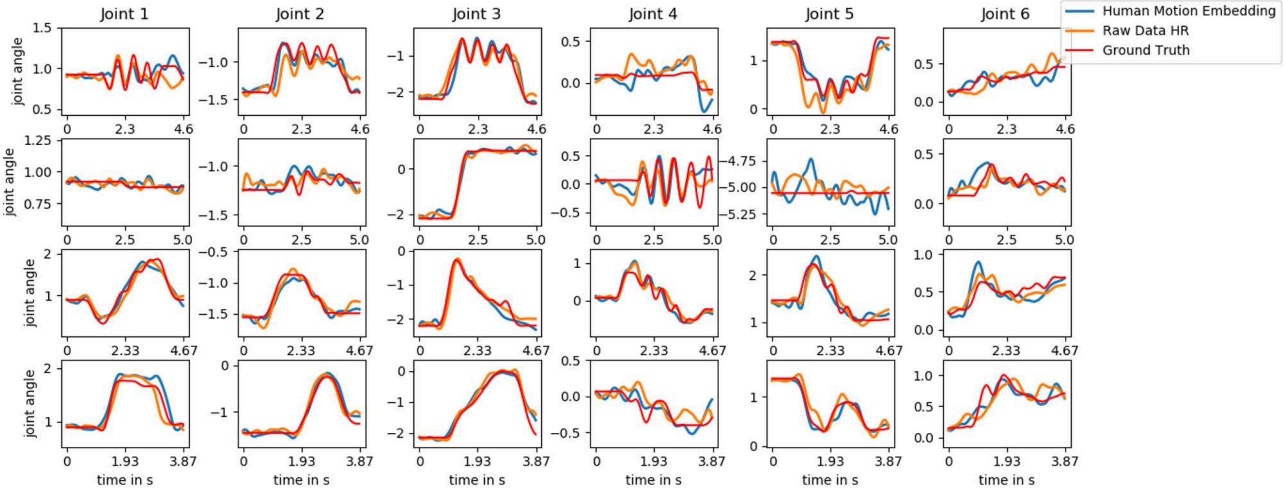


FIGURE 8 | The joint angle trajectory of joint 1–6 for a testing trial of each of the actions *hand-wave, hand-shake, parachute*, and *rocket* (top to bottom). We depict the ground truth data (red) compared to the predicted trajectory of the Human Motion Embedding model (blue) and the Raw Data HR model (orange). The predictions are performed as described in section 3.2 after initializing the models with ten observation steps. The Raw Data HR model is not able to generate human motion trajectories and is therefore provided with the last observed human pose.

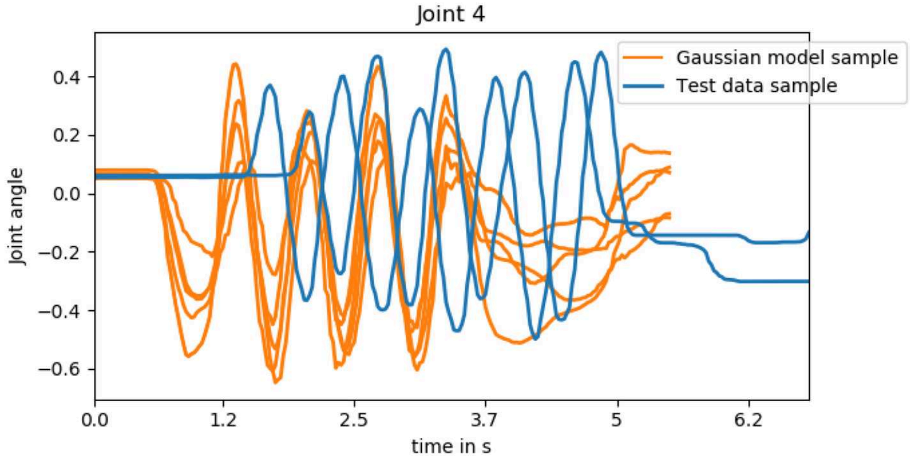


FIGURE 9 | Five samples from the Gaussian model of joint 4 of the hand-shake action and two testing data trials. As the Gaussian model is not adaptive, none of the samples is in accordance with any of the testing trials that are also depicted which is apparent in both the onset of motion and the phase shift.

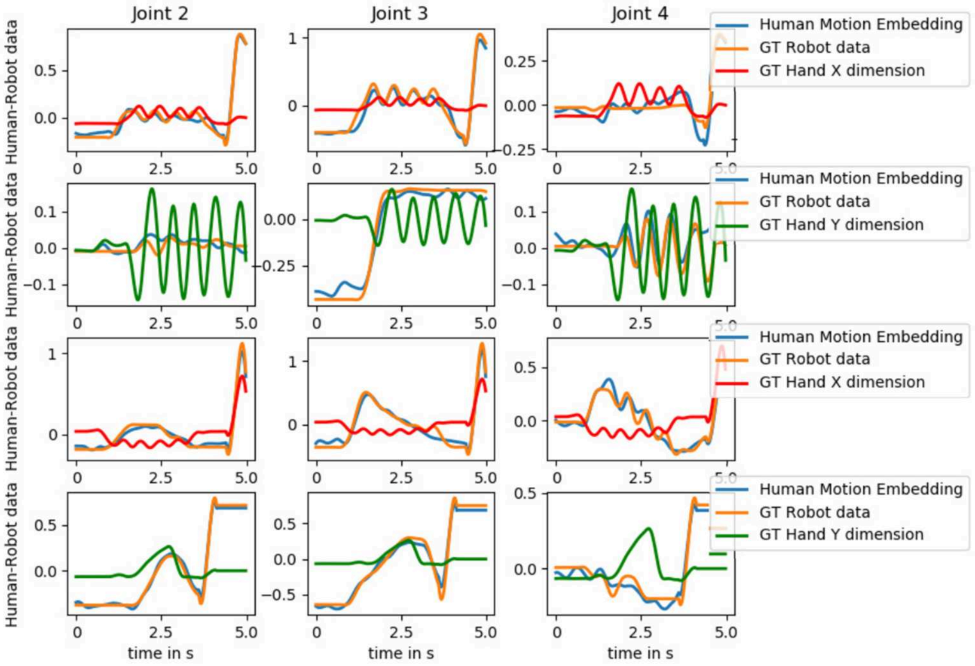


FIGURE 10 | The predictions of the Human Motion Embedding model as well as the ground truth robot motion data and the hand position (X or Y dimension) of the human for a testing trial of each interaction *hand-wave, hand-shake, parachute*, and *rocket* (top to bottom). The values are normalized to facilitate comparison. As not all joints are relevant to a task, we visualize joint 2–4. The predicted motion follows the oscillatory movement of the human hand during hand-wave (see joint 3), hand-shake (see joint 4), and the parachute (see joint 4) interaction as well the rise and fall of the rocket action (see joint 2 and 3).

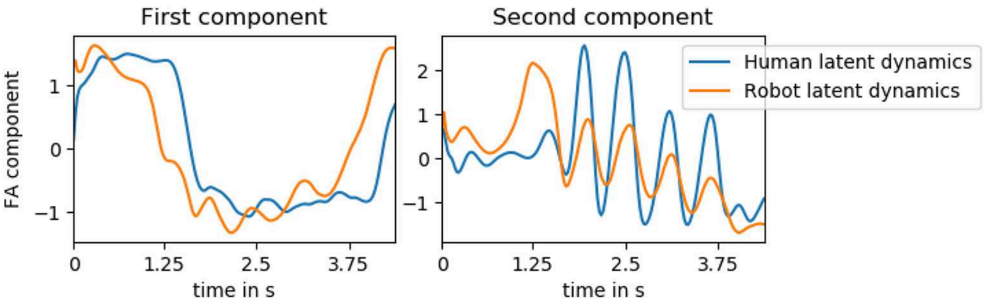


FIGURE 11 | The first two factor analysis (FA) components of a testing trial of the hand-shake interaction computed on both the latent variables extracted from the human ground truth motion and the latent variables predicting the robot motion. The first factor appears to represent the general onset, duration, and offset of the interaction while the second factor encodes the oscillatory motion of the hand and arm. Although the factor analysis is performed independently on the human and robot latent variables, the overall structure is similar. Additionally, the oscillatory motion is overlapping, indicating a coupling between the two systems.

coupling of the human partners during interaction, guide the generation of the robot which thus is coupled to its human partner.

After having established that the cheaper HHI data is required for high predictive performance (see section 5.1), we demonstrate that the extracted dynamics facilitate the performance of the predictive model of robot motion (see section 5.2). This indicates that the encoding of the future human motion and task dynamics can contribute to the robot's motion planning. This is in contrast to common approaches to imitation learning for interaction which use non-adaptive models. As we discuss in section 5.3, a non-adaptive trajectory model does not suffice in interactive tasks such as hand-shaking. With help of our generative approach, we can create synchronized behavior which shows a level of entrainment between human and robot (see section 5.4).

We believe that prediction and adaptation are essential to allow for natural HRI in shared workspaces. In future work, we plan to employ the system in real-time and to extend it to more complex tasks.

DATA AVAILABILITY STATEMENT

The datasets for this study can be found in the transfer learning for interaction repository, https://github.com/jbutepage/human_robot_interaction_data.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

JB contributed to the idea development and data collection, developed the methodology, implemented and trained the models, evaluated the models, and wrote the manuscript. AG contributed to the idea, data collection, and development of the robot software. ÖÖ contributed to the data collection and implementation of other baselines. MB and DK supervised the work.

FUNDING

This work was supported by the EU through the project socSMCs (H2020-FETPROACT-2014) and the Swedish Foundation for Strategic Research and EnTimeMent (H2020-FETPROACT-824160), and the Knut and Alice Wallenberg Foundation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frobt.2020.00047/full#supplementary-material

Supplementary Video 1 | The video demonstrates data collection for human-human and human-robot interaction as well as online employement of the trained models.

REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems.* Available online at: tensorflow.org.

Alissandrakis, A., Nehaniv, C. L., and Dautenhahn, K. (2007). Correspondence mapping induced state and action metrics for robotic imitation. *IEEE Trans. Syst. Man Cybernet. Part B* 37, 299–307. doi: 10.1109/TSMCB.2006.886947

Argall, B. D., Chernova, S., Veloso, M., and Browning, B. (2009). A survey of robot learning from demonstration. *Robot. Auton. Syst.* 57, 469–483. doi: 10.1016/j.robot.2008.10.024

Billard, A., Calinon, S., Dillmann, R., and Schaal, S. (2008). "Robot programming by demonstration," in *Springer Handbook of Robotics*, eds B. Siciliano and O. Khatib (Secaucus, NJ: Springer), 1371–1394.

Brownell, C. A. (2011). Early developments in joint action. *Rev. Philos. Psychol.* 2, 193–211. doi: 10.1007/s13164-011-0056-1

Bütepage, J., Kjellström, H., and Kragic, D. (2018a). "Anticipating many futures: online human motion prediction and generation for human-robot interaction," in *2018 IEEE International Conference on Robotics and Automation (ICRA)* (Brisbane, QLD: IEEE), 1–9. doi: 10.1109/ICRA.2018.8460651

Bütepage, J., Kjellström, H., and Kragic, D. (2018b). Classify, predict, detect, anticipate and synthesize: hierarchical recurrent latent variable models for human activity modeling. *arXiv[Preprint]. arXiv:1809.08875.*

Calinon, S., D'halluin, F., Sauser, E. L., Caldwell, D. G., and Billard, A. G. (2010). Learning and reproduction of gestures by imitation. *IEEE Robot. Autom. Mag.* 17, 44–54. doi: 10.1109/MRA.2010.936947

Dautenhahn, K. (2007). Socially intelligent robots: dimensions of human–robot interaction. *Philos. Trans. R. Soc. B Biol. Sci.* 362, 679–704. doi: 10.1098/rstb.2006.2004

Dinh, L., Sohl-Dickstein, J., and Bengio, S. (2017). "Density estimation using real NVP," in *International Conference on Learning Representations (ICLR)* (Toulon).

Dong, S., and Williams, B. (2011). Motion learning in variable environments using probabilistic flow tubes. *Int. J. Soc. Robot.* 4, 357–368. doi: 10.1109/ICRA.2011.5980530

Dong, S., and Williams, B. (2012). Learning and recognition of hybrid manipulation motions in variable environments using probabilistic flow tubes. *Int. J. Soc. Robot.* 4, 357–368. doi: 10.1007/s12369-012-0155-x

Ghadirzadeh, A., Maki, A., Kragic, D., and Björkman, M. (2017). "Deep predictive policy training using reinforcement learning," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Vancouver, BC: IEEE), 2351–2358. doi: 10.1109/IROS.2017.8206046

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). "Generative adversarial nets," in *NIPS* (Montreal, QC).

Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., et al. (2017). "beta-VAE: learning basic visual concepts with a constrained variational framework," in *International Conference on Learning Representations (ICLR)* (Toulon), 6.

Kingma, D. P., and Dhariwal, P. (2018). "Glow: generative flow with invertible 1x1 convolutions," in *Advances in Neural Information Processing Systems* (Montreal, QC), 10236–10245.

Kingma, D. P., and Welling, M. (2015). "Auto-encoding variational bayes," in *International Conference on Learning Representations (ICLR)* (San Diego, CA).

Koppula, H. S., and Saxena, A. (2015). Anticipating human activities using object affordances for reactive robotic response. *IEEE Trans. Pattern Anal. Mach. Intell.* 38, 14–29. doi: 10.1109/TPAMI.2015.2430335

Li, Y., Song, J., and Ermon, S. (2017). "Infogail: Interpretable imitation learning from visual demonstrations," in *Advances in Neural Information Processing Systems* (Montreal, QC), 3812–3822.

Maeda, G., Ewerton, M., Neumann, G., Lioutikov, R., and Peters, J. (2017a). Phase estimation for fast action recognition and trajectory generation in human–robot collaboration. *Int. J. Robot. Res.* 36, 1579–1594. doi: 10.1177/0278364917693927

Maeda, G. J., Neumann, G., Ewerton, M., Lioutikov, R., Kroemer, O., and Peters, J. (2017b). Probabilistic movement primitives for coordination of multiple human–robot collaborative tasks. *Auton. Robots* 41, 593–612. doi: 10.1007/s10514-016-9556-2

Osa, T., Pajarinen, J., Neumann, G., Bagnell, J. A., Abbeel, P., Peters, J., et al. (2018). An algorithmic perspective on imitation learning. *Found. Trends® Robot.* 7, 1–179. doi: 10.1561/2300000053

Rezende, D. J., Mohamed, S., and Wierstra, D. (2014). "Stochastic backpropagation and approximate inference in deep generative models," in *International Conference on Machine Learning* (Beijing), 1278–1286.

Rogoff, B., Mistry, J., Göncü, A., Mosier, C., Chavajay, P., and Heath, S. B. (1993). Guided participation in cultural activity by toddlers and caregivers. *Monogr. Soc. Res. Child Dev*. 58, v–vi, 1–174; discussion: 175–179. doi: 10.2307/1166109

Sakoe, H., and Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust. Speech Signal Proc*. 26, 43–49. doi: 10.1109/TASSP.1978.1163055

Sebanz, N., Bekkering, H., and Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends Cogn. Sci.* 10, 70–76. doi: 10.1016/j.tics.2005.12.009

Vesper, C., Abramova, E., Bütepage, J., Ciardo, F., Crossey, B., Effenberg, A., et al. (2017). Joint action: mental representations, shared information and general mechanisms for coordinating with others. *Front. Psychol.* 7:2039. doi: 10.3389/fpsyg.2016.02039

Zhang, C., Butepage, J., Kjellstrom, H., and Mandt, S. (2018). Advances in variational inference. *IEEE Trans. Pattern Anal. Mach. Intell*. 41, 2008–2026. doi: 10.1109/TPAMI.2018.2889774

Zhang, T., McCarthy, Z., Jow, O., Lee, D., Chen, X., Goldberg, K., et al. (2018). "Deep imitation learning for complex manipulation tasks from virtual reality teleoperation," in *2018 IEEE International Conference on Robotics and Automation (ICRA)* (IEEE), 1–8. doi: 10.1109/ICRA.2018.8461249

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Human Neuroscience

ORIGINAL RESEARCH
published: 25 May 2020
doi: 10.3389/fnhum.2020.00185

Tilting Together: An Information-Theoretic Characterization of Behavioral Roles in Rhythmic Dyadic Interaction

Dari Trendafilov [1†], Gerd Schmitz [2†], Tong-Hun Hwang [2], Alfred O. Effenberg [2] and Daniel Polani [3]*

[1] Institute of Pervasive Computing, Johannes Kepler University, Linz, Austria, [2] Department of Humanities, Institute of Sports Science, Leibniz University Hannover, Hanover, Germany, [3] Adaptive Systems, University of Hertfordshire, Hatfield, United Kingdom

Every joint collaborative physical activity performed by a group of people, e.g., carrying a table, typically leads to the emergence of spatiotemporal coordination of individual motor behavior. Such interpersonal coordination can arise solely based on the observation of the partners' and/or the object's movements, without the presence of verbal communication. In this paper, we investigate how the social coupling between two individuals in a collaborative task translates into measured objective and subjective performance indicators recorded in two different studies. We analyse the trends in the dyadic interrelationship based on the information-theoretic measure of transfer entropy and identify emerging leader-follower roles. In our experimental paradigm, the actions of the pair of subjects are continuously and seamlessly fused, resulting in a joint control of an object simulated on a tablet computer. Subjects need to synchronize their movements with a 90° phase difference in order to keep the object (a ball) rotating precisely on a predefined circular or elliptic trajectory on a tablet device. Results demonstrate how the identification of causal dependencies in this social interaction task could reveal specific trends in human behavior and provide insights into the emergence of social sensorimotor contingencies.

Keywords: sensorimotor contingencies, interpersonal coordination, collaborative interaction, transfer entropy, information theory, causality, social interaction

Edited by:
Leonhard Schilbach,
Ludwig Maximilian University of Munich, Germany

Reviewed by:
Guillaume Dumas,
Institut Pasteur, France
Tom Froese,
Okinawa Institute of Science and Technology Graduate University, Japan

****Correspondence:***
Dari Trendafilov
trendafilov@pervasive.jku.at

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience

Received: *17 December 2019*
Accepted: *27 April 2020*
Published: *25 May 2020*

Citation:
Trendafilov D, Schmitz G, Hwang T-H, Effenberg AO and Polani D (2020) Tilting Together: An Information-Theoretic Characterization of Behavioral Roles in Rhythmic Dyadic Interaction. Front. Hum. Neurosci. 14:185. doi: 10.3389/fnhum.2020.00185

1. INTRODUCTION

In everyday joint physical activities humans often coordinate their motor behavior. Such interpersonal coordination emerges when two people dance, row a canoe, or carry an object together. In some cases, coordination of this kind could be controlled through a direct physical contact (e.g., dance), and in other cases it could be mediated by a rigid object (e.g., a table), or it can also be distantly coordinated without any physical contact. In such various types of social interaction, visual contact has different levels of importance, as humans typically coordinate their movements by detecting visual movement information (Schmidt et al., 1990), and this could lead to coordination even when it is not necessary for completing the task (Schmidt and O'Brien, 1997; Richardson et al., 2005). From a dynamical systems perspective, such visually mediated interpersonal coordination can be understood as a self-organized entrainment process of biological rhythms (Newtson et al., 1987; Schmidt et al., 1990).

Interpersonal coordination can be influenced by different modes of non-verbal communication (e.g., mimicry, gestures, and facial expressions) as a basis of social interaction (Vicaria and Dickens, 2016). Such non-verbal expressions could induce spatiotemporal coordination and could facilitate social entrainment between two or more individuals (Phillips-Silver and Keller, 2012). Non-verbal means of communication are generally faster than verbal in sharing action plans and strategies, when instant reaction is required in a joint task (Knoblich and Jordan, 2003). Non-verbal communication modes, supporting emergent coordination, stretch across a broad spectrum of perceptual modalities, like visual, kinesthetic, tactile, or auditory (Marsh et al., 2009). Dancers coordinate non-verbally during performance relying on visual as well as auditory cues (Waterhouse et al., 2014). Demos et al. (2012) reasoned that the spontaneous coordination would result from emergent perceptuo-motor couplings in the brain (Kelso, 1995). Keller suggested that online perceptual information might enhance the anticipation of one's own action, as well as the co-performer's action, in terms of developing common predictive internal models (Keller and Appel, 2010; Keller, 2012).

One limitation of interpersonal coordination research stems from the fact that studies (e.g., Schmidt and Turvey, 1994) usually require individuals to focus their visual attention directly toward the movements of their co-actor. The current study tested the coupling strength and the stability of interpersonal coordination in a task that required visual control of a ball on a tablet screen. The movement of the ball resulted from the joint action of both persons. Neither the effect from the own action nor the partner's action could be perceived in isolation. When agents engage in social interaction, a rich spectrum of possibilities arises: under some conditions, they act together as one single entity, in other conditions they may act as independent individuals. There is an interplay between intrinsic, cognitively driven coordination and coordination driven by the environment. The intrinsic coordination between the actions of interacting agents is a candidate for a measure of individuality or autonomy with respect to other agents (Bertschinger et al., 2008). In a cooperative task, when two agents use independent controllers under information processing constraints, they arrive at intrinsic coordination in order to overcome limitations of their environment (Harder et al., 2010).

Interpersonal synergies are higher-order control systems formed by coupling movement of two (or more) actors. Many different approaches have been utilized for the characterization of social couplings, such as autocorrelation, cross-correlation (Box and Jenkins, 1970), transfer entropy (Barnett et al., 2009), Granger causality (Granger, 1969), and their potential has been demonstrated in many applications (e.g., Valdes-Sosa et al., 2005; Arnold et al., 2007; Ryali et al., 2011). Interactive alignment was used to investigate interpersonal synergies in conversational dialog (Fusaroli et al., 2014; Fusaroli and Tylén, 2015). A key challenge is to design a suitable procedure that allows synchrony and turn-taking to spontaneously take place. Traditional interactive paradigms mainly consist of non-contingent social stimuli that do not allow true social interaction (Redcay et al., 2010). However, apparent interpersonal coordination could be merely incidental rather than reflecting true coordination—people may appear to coordinate their movements because they simultaneously execute similar motor programs, mediated by shared motor representations (Garrod and Pickering, 2004, 2009; Sebanz et al., 2006). In this study we addressed that by designing a performance oriented closed-loop interaction paradigm, which requires tightly-coupled motor coordination. A study, based on the perceptual crossing paradigm, investigates the direction of influence using discretized turn-taking events (Kojima et al., 2017).

Dynamical processes modeling the stable modes of intentional inter-limb coordination within (Haken et al., 1985) and between (Schmidt et al., 1998) individuals, can be represented by coupled oscillators. One of the main principled treatments of mutual synchronization in a network of oscillators was proposed by Kuramoto (1984) and is related to work of Bottani (1996), Pikovsky et al. (2001), Strogatz (2003), and Winfree (1967, 1980). Kuramoto (1984) developed a tractable mean-field model of coupled biological oscillators (Winfree, 1967), such as groups of chorusing crickets (Walker, 1969), flashing fireflies (Buck, 1988), or cardiac pacemaker cells (Peskin, 1975), which exhibits a spontaneous transition from incoherence to collective synchronization as the coupling strength is increased past a certain threshold. However, the original model relates to sinusoidal all-to-all couplings, which are not typical for biological systems. Strogatz (2003) introduced a not pure sinusoidal generalization, which also enables the addition of noise by a flux term. In the case of identical oscillators, perfect synchrony extends to time-delayed interactions, and when the oscillators are completely disorganized, different synchronized states can coexist with a stable incoherent state (Adlakha et al., 2012). Hanson's model of firefly entrainment, captured by an extension of the Haken-Kelso-Bunz equation (Kelso et al., 1990), specifies the eigenfrequency difference or frequency detuning between two rhythmic units. It reveals that human interpersonal rhythmic coordination is subject to the same dynamical laws as seen elsewhere in nature. Entrainment of unpredictable and chaotic systems was studied more recently by Dotov and Froese (2018).

New approaches from social neuroscience use imaging techniques, such as FMRI, fNIRS, and M/EEG, to study brain mechanisms in social interactions. One promising approach is hyperscanning, in which the brain dynamics of multiple subjects are studied simultaneously (Czeszumski et al., 2020). With EEG-hyperscanning, Sänger et al. (2012) found increased phase locking and phase coherence connection strengths in phases characterized by high demands on (musical) action coordination. Furthermore, oscillatory couplings between musicians' brains enabled the inference of leader-follower roles (Sänger et al., 2013). Similar observations were made by Dumas et al. (2010) in an imitation task, i.e., neuronal synchronization becomes asymmetric when one person is a leader and the other imitator. Konvalinka et al. (2014) demonstrated that multivariate decoding of inter-brain activity in an interactive task can identify the spontaneous emergence of leader-follower relationships within a dyad. Stephens and Galloway (2017) applied a quantitative information-theoretic approach for modeling the information exchange in healthcare teams in interactive navigation by transforming EEG-data into a stream of Shannon-entropy units characterizing team members' relationships.

Studies focusing on hyperscanning analysis of information flows between human brains require estimating the causal links between brains. Such causal links are established typically using Granger Causality or its frequency domain equivalent Partial Directed Coherence (PDC) (e.g., Astolfi et al., 2011, 2012). Previous results reveal stronger causal links during increased cooperative behavior and altruistic behaviors in decision-making tasks (Fallani et al., 2010; Ciaramidaro et al., 2018). Schippers et al. (2010) studied causal links in gesture communication using fMRI and Pan et al. (2017) using fNIRS between brains of cooperating lovers. Yun et al. (2012) investigated a paradigm for identifying the behavioral and the neural correlates of implicit cooperative social interaction. Leong et al. (2017) demonstrated that adults and infants show significant mutual neural coupling during social interactions. Liu et al. (2016) proposed a novel method for studying social cognition in the cooperative and obstructive game of Jenga. Naeem et al. (2012) explored mutual information on EEG data in social interaction tasks. Lobier et al. (2013) found that Phase Transfer Entropy detects the strength and direction of connectivity in the presence of noise characteristic for EEG data. The growing variety of hyperscanning analysis techniques suggest their exploratory nature and often the advantages and disadvantages of a specific method are not obvious. A key open research question relates to the neural substrates enabling the information flow between brains. In this respect it is crucial to emphasize the difference between information flow and synchronized neural activity between brains due to identical sensory input.

Despite such significant insights into the neuronal mechanisms of social interactions and social roles, Liu et al. (2018) pointed out that their behavioral correlates are still largely unclear and further research is needed to decompose the complicated mental constituent into basic psychological processes. The reciprocal influence in social interactions represents a major challenge with regard to the design of experiments. This is a starting point for the present study, which introduces a behavioral approach to quantify and investigate reciprocal influences and social roles. In a two-person cooperative tapping behavioral study using transfer entropy Takamizawa and Kawasaki (2019) identified leader/follower relationships which were consistent with subjective experiences.

2. MEASURES OF CAUSAL RELATIONSHIP

Various measures of causal relationship exist, the main groups being model-based [e.g., Granger causality (Granger, 1969) or dynamic causal modeling (Friston et al., 2003)] or non-parametric methods [e.g., transfer entropy (Schreiber, 2000) or directed information (Massey, 1990)]. Granger causality is particularly useful when the interaction between the agents can be approximated well linearly and data has relatively low levels of noise (Nalatore et al., 2007).

Shannon mutual information, in conjunction with signal independent component analysis provides new aspects of brain-to-brain coupling in dyadic social interactions (Naeem et al., 2012), and reveals how the dynamic interaction unfold, determined by its specific properties. In the context of information theory, the key measure of information of a random variable is its Shannon entropy (Shannon, 1948). The entropy quantifies the reduction of uncertainty obtained when one actually measures the value of the variable. Therefore, if prediction enhancement can be associated to uncertainty reduction, it is expected that a causality measure would be naturally expressible in terms of information-theoretic concepts. Attempts to obtain model-free measures of the relationship between two random variables based on mutual information (MI) do not rely on any specific model of the data. However, MI says little about causal relationships, because of its lack of directional and dynamical information. Since MI is symmetric under the exchange of signals, it cannot distinguish driver and response systems, and furthermore, standard MI only captures the amount of information shared by two signals. In contrast, a causal dependence is related to the information being exchanged, rather than shared. The principle of maximum causal entropy provides causal analysis of the behavior of interacting systems, reflecting the causal dependencies between the processes (Ziebart, 2013; Ziebart et al., 2013). Building upon Massey's directed information (Massey, 1990) it extends random field models to settings with feedback, by providing a framework for estimating an unknown process based on its interactions with a known process.

Another information-theoretic framework, called transfer entropy, was proposed by Schreiber (2000) as a rigorous derivation of a Wiener causal measure. Assuming that two time series of interest $X = x_t$ and $Y = y_t$ can be approximated by Markov processes, transfer entropy computes the deviation from the following generalized Markov condition

$$p(y_{t+1}|y_t^n, x_t^m) = p(y_{t+1}|y_t^n), \quad (1)$$

where $x_t^m = (x_t, ..., x_{t-m+1}), y_t^n = (y_t, ..., y_{t-n+1})$, and m and n are the orders (memory) of the Markov processes X and Y, respectively. Using the expected Kullback-Leibler divergence between the two probability distributions at each side of Equation 1, defines transfer entropy from X to Y as

$$TE(X \rightarrow Y) = \sum_{y_{t+1}, y_t^n, x_t^m} p(y_{t+1}, y_t^n, x_t^m) \log \frac{p(y_{t+1}|y_t^n, x_t^m)}{p(y_{t+1}|y_t^n)}. \quad (2)$$

Transfer entropy naturally incorporates directional and dynamical information, because it is inherently asymmetric and based on transition probabilities. Earlier efforts to understand causal relationships mostly relied on model-based approaches, such as Granger causality or dynamic causal modeling. In contrast, transfer entropy (TE) does not require a model of the interaction and is inherently non-linear. Thus, the sensitivity of transfer entropy to all order correlations becomes an advantage for exploratory analyses over Granger causality or other model-based approaches. This is particularly relevant when the detection of unknown non-linear interactions is required. Transfer entropy has seen a dramatic surge of interest in neuroscience recently, where it is used to estimate the information transfer between two tightly coupled processes. It requires the observation of multiple realizations of the processes, in order to estimate the associated probability density functions, provided stationarity assumptions. In this study, we investigated the applicability of TE as a measure characterizing causal dependence based on behavioral data in a simple collaborative motor task and demonstrated the relation of TE to standard performance metrics.

3. THE STUDY

The current study builds on the tetherball paradigm introduced in Hwang et al. (2018), and is implemented on a tablet computer (see **Figure 1**). With rhythmic finger movements, a pair of participants had to tilt the tablet in order to rotate a ball along a predefined circular target trajectory (experiment A). One measure of joint task performance in this scenario is the average target tracking precision, i.e., the spatial error between the ball and the target trajectory. Since Hwang et al. (2018) reported for the circular task that the error reaches a plateau after a few trials, we applied an alternative task in experiment B, where the participants had to track a rotating ellipse instead. In each condition, we evaluated the tracking accuracy as a measure of task performance as well as the information flow (i.e., transfer entropy) as a measure of mutual influence between the two participants based on their actions. Participants were also asked to report on their subjective experience of interpersonal coordination. Our aim was to gain an initial insight into the utility of information-theoretic functionals, such as transfer entropy and its variants, for the characterization of social couplings based on behavioral data. We investigated the following research questions:

- (RQ1) What is the relation between measured interpersonal coordination and achieved task performance?
- (RQ2) Is there a correlation between objective and subjective measures of interpersonal coordination?
- (RQ3) Can transfer entropy provide insights into specific behavioral patterns and identify distinct roles within a dyad?

4. MATERIALS AND METHODS

4.1. Participants

We collected data from 76 participants (provided in the **Supplementary Material**) and reanalyzed data of the 72 participants from the study of Hwang et al. (2018). In total, the data from 46 females and 102 males were analyzed (mean age: 25.7 years, *SD*: 4.6 years). All participants reported to healthy, and none of them had overt psychic or cognitive impairments. They were tested for normal eyesight with the Landolt rings chart (Jochen Meyer-Hilberg) and for normal hearing abilities with the HTTS audio test (SAX GmbH). All participants gave their written informed consent to the study. The study protocol and all documents had been independently reviewed and pre-approved by the Ethics Committee of the Leibniz University Hannover.

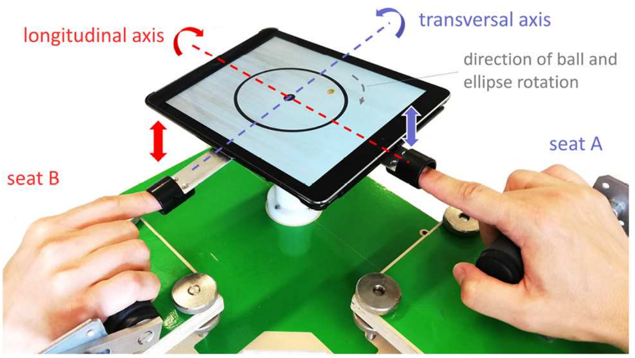


FIGURE 1 | Experimental apparatus. Two participants moved a virtual ball on a circular or elliptic target line. A universal joint underneath the tablet limited the motion space to motions around the longitudinal and the transversal axes. Each participant controlled one tablet axis with upward and downward movements of his index finger tip. Therefore, the task required the coordination of the individual actions.

4.2. Experimental Apparatus

The participants sat in front of the experimental apparatus, which is shown in **Figure 1**. With their dominant hand, they grabbed an adjustable handle and inserted the tip of their index finger into a lever, which was attached to a tablet (iPad Air, Apple Inc.). A universal joint underneath the tablet allowed to move the tablet around its transversal and longitudinal axes as indicated by the dashed lines in **Figure 1**. Rotations along the vertical axis were not possible. With upward and downward movements of the index finger, each participant controlled one dimension; i.e., participant A on seat A controlled the transversal axis and participant B on seat B controlled the longitudinal axis. The screen (1,024 * 768 pixel, 60 Hz) displayed a target line and a virtual ball. The target line had a width of 0.29 cm and had either the shape of a circle (diameter: 8.95 cm, experiment A) or the shape of an ellipse (axes length: 10.94 and 8.47 cm, experiment B). The ellipse rotated with 2.5 revolutions per minute. The ball is illustrated as gray dot in **Figure 1**. It had a diameter of 0.58 cm and was connected by an invisible elastic spring to an anchor at the center of the screen. The spring force was just strong enough to pull the ball to the anchor, when the tablet was in a horizontal position.

4.3. Paradigm

By tilting the tablet, the participants could move the ball around the central anchor. The instruction was to move the ball clockwise and as accurately as possible on the target line. This was only possible if both players contributed to the task and tilted the tablet around both axes in a certain pattern and with a certain amplitude of frequency. Since both players sat in an angle of 90° to each other, optimal performance was achieved by synchronizing finger movements with a 90° phase-difference [see Video 1 in the Supplementary Material of the earlier publication (Hwang et al., 2018)].

Two participants of the same gender performed as dyad. Twenty-eight female and 22 male dyads performed the task with a circle as target line (experiment A) and 24 dyads performed the task with an ellipse as target line (experiment B). Data from 36 dyads from experiment A had been published before in Hwang et al. (2018). These authors focused on a different topic compared to the present study by investigating the impact of different types of auditory feedback on joint performance in the tablet task. Thereto, they compared the performance between four groups, which played under different perceptual conditions. One group received purely visual information and three groups visual plus auditory information. The auditory information either provided knowledge of performance by transforming the angular velocity measured by the tablet's gyroscopes into sound (broom sweeping sound) or knowledge of results by transforming the two-dimensional ball position on the screen sound (synthesized violin, details of the parameter-sound-transformation are described in Hwang et al., 2018). Yet unpublished data from further 14 dyads with a different tilt sound (synthesized violin) but similar performance compared to the participants from Hwang et al. (2018) were also included in the analyses. We believe that reanalyzing these data sets is reasonable considering the different study goals and types of analyses: Whereas Hwang et al. (2018) focused on comparisons between groups, the present study focused on the intra-dyadic coupling between two players and leader-follower relationships. An influence of sound condition on leader-follower-relationships was not expected, because both players of one dyad had the same perceptual condition. Thus, we combined the data of all groups for the analyses of the present study. Nevertheless, we analyzed whether sound condition influences social coupling by comparing the transfer entropy measures between groups. The participants from experiment B were not provided with artificial auditory information. This paradigm was specifically designed to investigate leader-follower relationships, for which we explored transfer entropy as an objective measure of causal dependence. This could serve as an initial step toward the characterization of more general social sensorimotor contingencies.

4.4. Procedure

The participants familiarized with the experimental apparatus in a 2-min practice phase. During that, each participant controlled his own ball and tried to track a target zone (diameter 0.58 cm), which moved randomly along the longitudinal axis for participant A and along the transversal axis for participant B. The main task was divided up into 1-min trials. Because the task was attentionally demanding, a 2-min break was introduced after every five trials. During that break, the participants were allowed to talk with each other, but instructed not to talk about the experiment. The participants from experiment A performed fifteen trials. In experiment B, the participants performed thirty trials. In the latter group, the participants exchanged their seats after every five trials; i.e., each participant performed fifteen trials in seat A and fifteen trials in seat B. This procedure was chosen, because the 90° angles of the seating positions and the resulting 90° phase-difference of the player's actions might influence the leader-follower relationship: The ball first passes player A, then player B; therefore, player A might take the leadership role more likely as player B. By exchanging seating positions, we could analyze leader follower relationships independently from this effect. To assess subjective experiences, the participants were interviewed at the end of the experiment using standardized questions. In an open question, the participants were asked to describe who they felt was leading the interaction, if any at all. Furthermore, participants had to rate on a 7-point Likert scale how much they felt they helped their partner (Q1) and how much they felt their partners helped them (Q2).

4.5. Dependent Measures

The tablet recorded the path of the ball (from the visual display) and the angular velocity (from the built-in gyroscope sensor) at the sampling rate of 60 Hz. We measured task-related performance based on the absolute error between the target trajectory and the actual ball trajectory. Furthermore, we computed the transfer entropy between the two players' actions per trial, using the Kraskov-Stogbauer-Grassberger transfer entropy estimator (Kraskov et al., 2004) from the JIDT toolkit (Lizier, 2014), based on the raw tablet gyroscope time series for the transversal and longitudinal axes, which correspond to the finger movements of players A and B, respectively. We computed the transfer entropy and the mean levels of the absolute error over each 1-min trial while discarding the initial 8.3 s (500 samples at 60 Hz) in every trial for initialization reasons.

4.6. Data Reduction

In order to relate the user ratings of perceived collaboration to the objective levels of transfer entropy we had to perform specific conversion on the subjective data. We transcribed the informal verbal answers from both players with discrete numerical representations $\{1, 0, -1\}$, meaning, respectively {I was leader, there was no leader at all, partner was leader}. Using this numerical representation, we subtracted the values in order to compute the difference in the opinions and took the sign of the result. Furthermore, for the closed questions, we calculated the difference between the ratings of the partners and took the sign of the result.

5. RESULTS

In order to validate the transfer entropy measure for this particular data set, we performed surrogate data testing with 1,000 random pairings for each TE value in each trial. The results of the surrogate data testing are shown in **Figure 2**. Using Wilcoxon tests on the surrogate data, we compared the outcome of each trial against the constant 0.05, which corresponds to the conventional significance level. The TE values were significant in all trials of experiment A (**Figure 2A**) (at least p <0.05) and in most trials of experiment B (**Figure 2B**, p >0.05 in trials 1, 4, 6, and 28, at least p <0.05 in all other trials; results were Bonferroni-Holm-corrected). Transfer entropy estimates typically stabilized at 1,000 samples, with some variability across trials and subjects. However, such trends depend on sampling rates and on the performance of subjects. For our analysis, we computed TE on a trial base, i.e., ca. 3,000 data points.

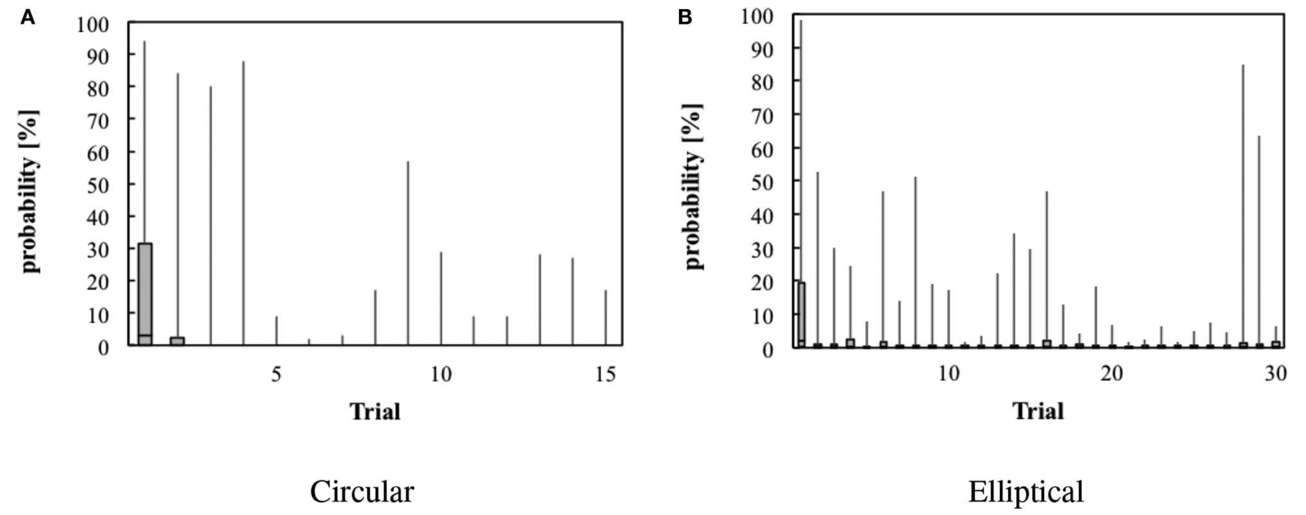


FIGURE 2 | Box and whisker plots illustrating the probability that the transfer entropy measures are random. Data of experiment A (circular target) are shown in **(A)**, data of experiment B in **(B)**. **(A)** Circular. **(B)** Elliptical.

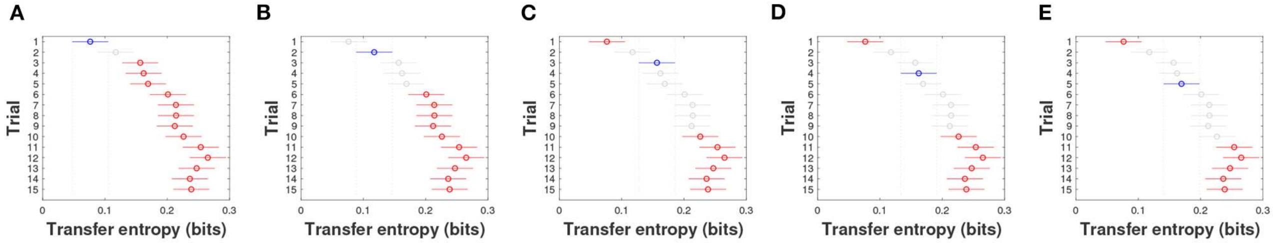


FIGURE 3 | Mean total transfer entropy levels ($TE(A \rightarrow B) + TE(B \rightarrow A)$) per trial averaged over all 50 pairs in Experiment A, revealing the learning effect and the statistical significant differences of the transfer entropy means between trials one to five in blue and the trials in red (in separate subplots). The differences were not significant for the rest of the trials. **(A)** Trial 1 vs. rest. **(B)** Trial 2 vs. rest. **(C)** Trial 3 vs. rest. **(D)** Trial 4 vs. rest. **(E)** Trial 5 vs. rest.

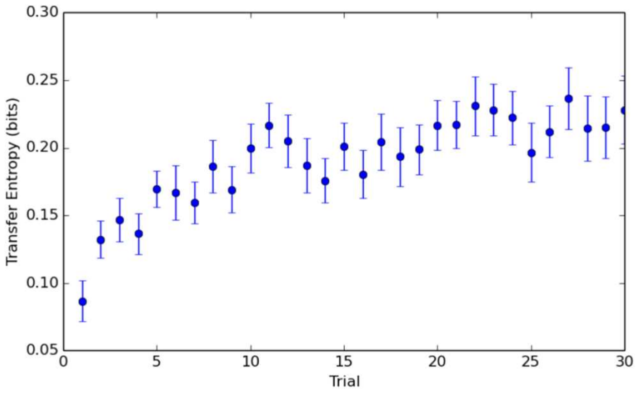


FIGURE 4 | Mean total transfer entropy levels ($TE(A \rightarrow B) + TE(B \rightarrow A)$) per trial averaged over all 24 pairs in Experiment B, revealing the learning effect.

5.1. Learning Effect

The trend in mean total transfer entropy levels averaged over all pairs, reveals a pronounced learning effect during both experiments (see **Figures 3**, **4**). Accordingly, trials were significantly different (experiment A: $F(14, 630) = 38.81$, $p < 0.001$; Experiment B: $F(29, 667) = 5,09$, $p < 0.001$). With the circular target, the trend is more consistent and with lower variance than with the elliptic target, presumably due to repeated seat exchanges in the latter. **Figure 3** reveals the statistical significant differences of transfer entropy means between trials one to five in blue and the trials in red in Experiment A, computed with Tukey-Kramer (HSD) multiple comparison test. The differences were not significant for the rest of the trials. These figures suggest that Experiment A and Experiment B, although slightly different by design, are TE-invariant and reach plateau at ∼0.25 bits.

Neither in Experiment A nor in Experiment B, seat position had a significant effect on the size of transfer entropy. Within experiment A, there was a significant difference between groups with different auditory conditions [$F_{(4,45)} = 2.85$, $p = 0.035$]: On average, transfer-entropy was higher for the participants that heard a broom sweeping sound (0.24, *SD*: 0.07 bits) compared to participants that heard a synthesized violin sound of the tablet tilt velocity (0.16, *SD*: 0.05 bits, $p = 0.024$).

5.2. Social Roles and TE Relevance

In order to get further insight into the interpersonal dynamics of the emerging social interaction, we computed the differences in transfer entropy ($TE(A \rightarrow B) - TE(B \rightarrow A)$ and $TE(B \rightarrow A) - TE(A \rightarrow B)$) between both players for each 1 min trial while taking into consideration the actual seating. The distribution of results on a trial level for all pairs from Experiment A suggests potential leader-follower roles for particular pairs despite the high variability (e.g., pairs 3, 4, 6, 9, 15, 16, 23, 24, 26, 27, 38, 40, 49) (see **Figure 5**). Predominantly positive values reflect consistently higher information transfer from player A to player B and vice versa, negative levels—from player B to player A, which provides a base for making specific inferences about leader-follower relationships spontaneously emerging during the non-verbal social interaction.

In order to compensate for the alternating seating arrangements in experiment B, we split the results into two subsets per pair (shown in blue and cyan in **Figure 6**), corresponding to the consistent seating of both players during the experiment. Blue color denotes trials in which player A sat in seat A and cyan in which player A sat in seat B. The figure reveals how repetitive seat exchanges affect the coordination trends. For example, for pair 9, the results show that player A transfers more entropy than player B regardless of the seat, reflected by positive and negative transfer entropy differences in the two different seating arrangements, which suggests consistent roles for this pair throughout the experiment. Similar trends are visible also for pairs 8, 10, 18, and 24, suggesting that their social roles were not affected by the particular seating. Identifying such coherent patterns of social behavior was one of the main goals of this study and in the next paragraph we demonstrate how these objective measures correspond to the subjective levels of interpersonal coordination measured by user experience questionnaires.

The participants rated both the support for (Q1) and from their partners (Q2) as high [Experiment A: medians = 6, interquartile ranges (IQRs) = 1; experiment B: medians = 5, IQR = 1 (Q1) and 2 (Q2)]. Furthermore, 18 percent (Experiment A) and 19 percent (Experiment B) of the participants answered that they were the leader, while 9 percent (Experiment A) and 17 percent (Experiment B) saw their partner as leader. Neither in Experiment A nor in Experiment B did these variables differ significantly between both players. There was no significant correlation between the open ended questions and Q1 or Q2 in either of the experiments. However, we found a significant correlation (Rho = 0.42, p <0.00002, 95% CI [0.24, 0.57]) between Q1 and Q2 in Experiment A. Since in both studies subjective user experience was collected via questionnaires only at the end of the experiment, we were not able to correlate the user ratings with transfer entropy levels on a trial base. Instead, we took the last third of the trials for each pair and considered the reduced subsets representative for the subjective ratings provided at the end. This assumption essentially takes into account both the learning and the memory effect. Furthermore, since the main goal of this study was to identify specific leader-follower relationships based on the direction of influence between the two players and not on the exact values, which are only important for providing the order of magnitude, we took the sign functions of both the transfer entropy differences and the user ratings differences. Following this approach, we computed the Pearson correlation between the transfer entropy differences of the last third of the trials and the user experience differences using their sign functions and found a significant correlation (Rho = 0.34, p <0.02, 95% CI [0.07, 0.57]) for the open ended question in experiment A (see **Figure 7A**). Furthermore, we found a significant correlation (Rho = −0.44, p <0.04, 95% CI [−0.7, −0.04]) for Q2 in experiment B (see **Figure 7B**). The negative correlation resonates well with the content of Q2, as Q1 and Q2 have opposite meanings in inferring a potential leadership. Here, we assume that stronger sense of leadership is associated with higher ratings of one own's influence (or help) and lower ratings of the other's influence (or help).

5.3. Leader-Follower vs. Performance Trends

To analyze the relationship between achieved task performance and objective as well as subjective measures of coordination, we first performed Bayesian linear regression analyses. Values deviating more than two standard deviations from the group mean were excluded from the analyses. As task performance is represented by the mean error, it was chosen as criterion variable. Predictors were $TE(A \rightarrow B)$, $TE(B \rightarrow A)$ and the subjective data from the questionnaires. In experiment A, the Bayes factor ($BF10 = 20.50$, percentage error < 0.001) was largest for a model including $TE(A \rightarrow B)$ as predictor [$R2 = 0.20, F_{(1,\,45)} = 11.20, p = 0.002$]; i.e., the data were 20.5 times more likely under this model than under the null model. In experiment B, the largest Bays factor was achieved for a model with the predictors $TE(A \rightarrow B)$ and the coded answer from participant A to the open question [$BF10 = 3.83$, percentage error < 0.01; $R2 = 3.83, F_{(2,\,21)} = 4.99, p = 0.017$]. **Figure 8** shows alternative models sorted by their Bayes factor. The analyses of both experiments suggest that among the tested variables measures from participant A are the most important predictors for the joint performance—despite the seat exchange in experiment B.

Another interesting observation is the relation between the transfer entropy differences and the normalized mean absolute error, which is shown in **Figure 9A** for experiment A and in **Figure 9B** for experiment B. In both studies, the point densities have a characteristic bell-shaped form, reflecting that low performance is associated with low levels of transfer entropy differences. This suggests that in cases of (i) low $TE(A \rightarrow B)$ and low $TE(B \rightarrow A)$ or (ii) quasi equal transfer entropies ($TE(A \rightarrow B) - TE(B \rightarrow A) \approx 0$), performance can be high or low alike, however, for more significant and disparate levels of transfer entropies performance is typically higher. This provides an interesting insight into the social aspect of the experimental paradigm, i.e., more pronounced behavioral roles of leader-follower eventually lead to higher performance, although such behavior is not necessary, as high performance could be achieved even with less-structured or more-balanced behavior from both partners. The results also suggest that the achieved performance per level of coordination was higher for the circular than for the elliptical target, revealed by lower absolute error yielded at similar levels of interpersonal coordination. This confirms, as expected,

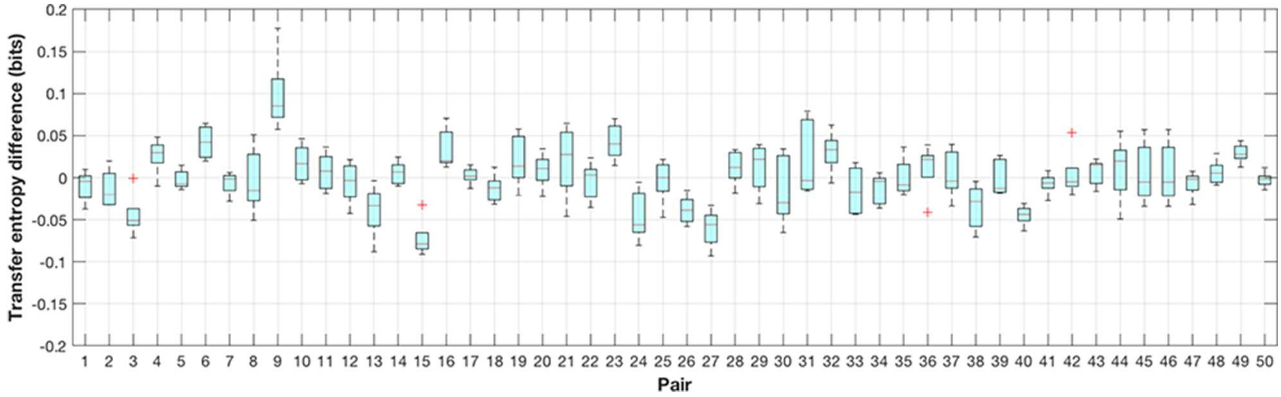


FIGURE 5 | Distribution of transfer entropy differences ($TE(A \rightarrow B) - TE(B \rightarrow A)$) computed on each 1 min trial for all pairs in experiment A (circular target). Consistent positive levels reveal higher information transfer from player A to player B and vice versa, negative levels—from player B to player A, which suggests specific social dynamics (e.g., pairs 3, 4, 6, 9, 15, 16, 23, 24, 26, 27, 38, 40, 49). Data for pairs from 1 to 36 is from an earlier study (Hwang et al., 2018).

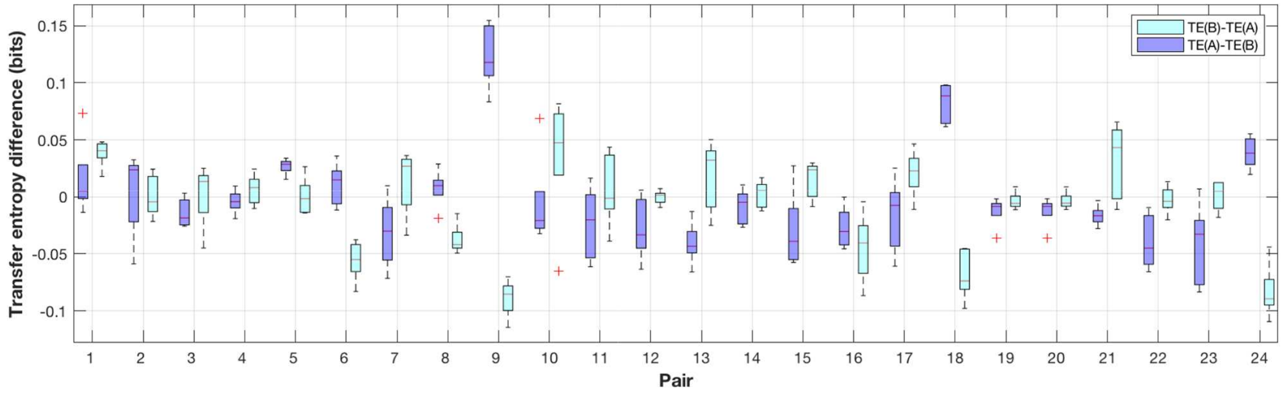


FIGURE 6 | Distribution of transfer entropy differences ($TE(A \rightarrow B) - TE(B \rightarrow A)$ and $TE(B \rightarrow A) - TE(A \rightarrow B)$) on a trial level for all pairs in experiment B (elliptical target). The results are split into two subsets per pair (blue/cyan), corresponding to consistent seating of players over different trials. Pairs 8, 9, 10, 18, and 24 exhibit opposite trends in the two seating arrangements, suggesting that their social roles were not affected by the particular seating.

that the more difficult elliptic task requires higher degree of social coordination between participants.

6. DISCUSSION

This study investigated the social dynamics of interpersonal coordination in two proximal collaborative experiments. In the tetherball paradigm, participant pairs were asked to tilt the tablet together for the task. We compared the mean levels of the error, transfer entropy and subjective ratings in our analysis. We hypothesized that stable rhythmic patterns of coordination would emerge in the course of interaction, which would be measurable with information-theoretic functionals. Our aim was to quantitatively identify and explain observed trends in the social aspect of interpersonal interaction. To test these predictions, we analyzed two types of studies, which examined the movement patterns of pairs of individuals performing a collaborative circular and elliptic motion jointly through the coordination of their movements. In both studies, the participants were only instructed to maximize their task performance, without explicitly guiding them to focus on their interpersonal entrainment and coordination. The topic of coordination was raised only in the user experience questionnaire, filled in after the experiment was completed. This ensured that the social dynamics, observed during the experiment, emerged spontaneously and not by instruction. The results presented in **Figures 8**, **9**, provide the answer to RQ1, which is invariant for both studies. Correlations, provided in **Figure 7** reflect RQ2, although the significance is rather sporadic and not across the board. **Figures 5**, **6** provide evidence supporting a positive answer to RQ3 in respect to both studies. The transfer entropy measure clearly emphasizes the learning effect, although the trend is not continuously monotonic. However, considering the fact that the relationship between task performance and interpersonal coordination (as measured by transfer entropy) is not monotonic, such a trend is plausible. The results show that tightly coupled interaction leading to

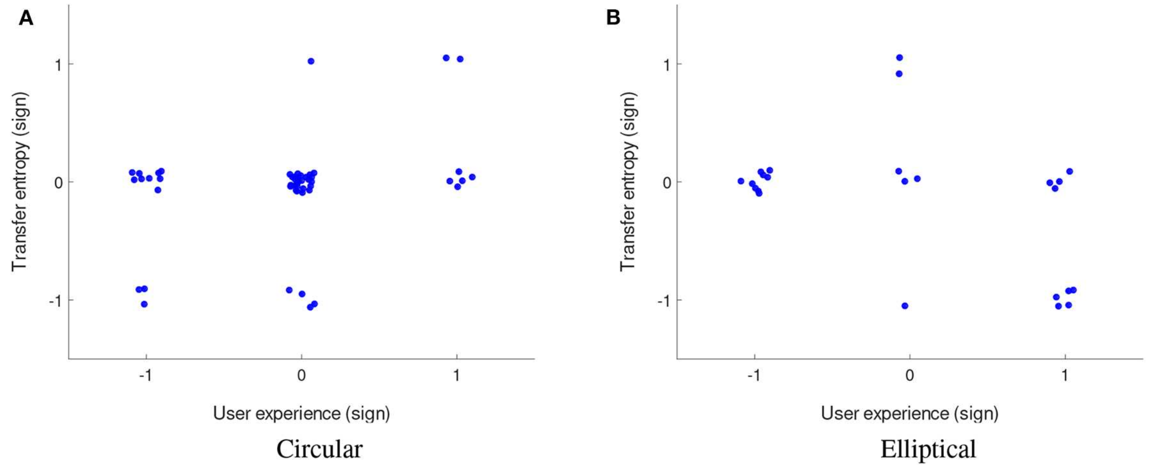
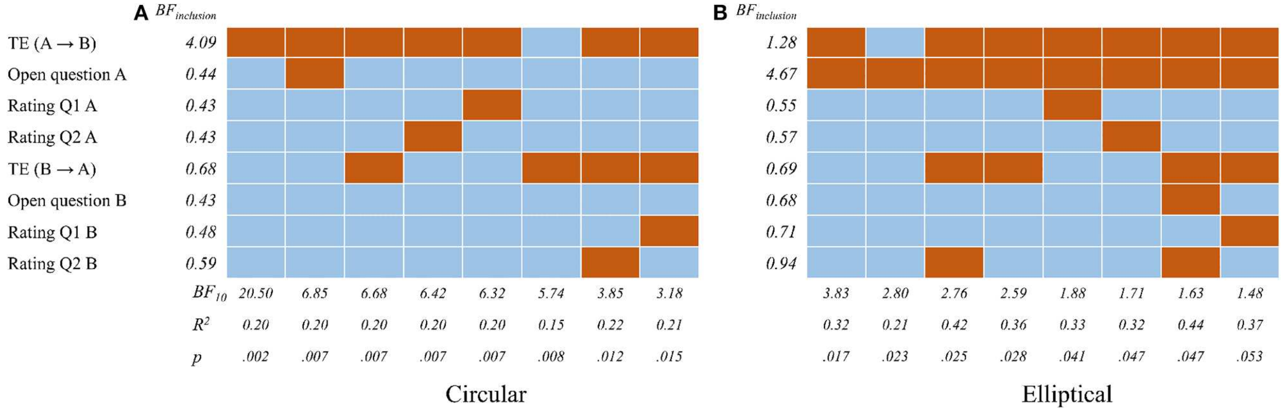


FIGURE 7 | Significant Pearson correlation between the sign functions of the transfer entropy difference and user experience difference for the open ended question in experiment A (circular target, **A**) (Rho = 0.34, *p* <0.02) and for Q2 in experiment B (elliptical target, **B**) (Rho = −0.44, *p* <0.04). To avoid overlapping points and improve visibility, the values (−1, 0, 1) of the sign functions are perturbed with white noise.



A

	$BF_{inclusion}$								
TE (A → B)	4.09	🟧	🟧	🟧	🟧	🟧	🟦	🟧	🟧
Open question A	0.44	🟦	🟧	🟦	🟦	🟦	🟦	🟦	🟦
Rating Q1 A	0.43	🟦	🟦	🟦	🟦	🟧	🟦	🟦	🟦
Rating Q2 A	0.43	🟦	🟦	🟦	🟧	🟦	🟦	🟦	🟦
TE (B → A)	0.68	🟦	🟦	🟧	🟦	🟦	🟧	🟧	🟧
Open question B	0.43	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟦
Rating Q1 B	0.48	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟧
Rating Q2 B	0.59	🟦	🟦	🟦	🟦	🟦	🟦	🟧	🟦
	BF_{10}	20.50	6.85	6.68	6.42	6.32	5.74	3.85	3.18
	R^2	0.20	0.20	0.20	0.20	0.20	0.15	0.22	0.21
	p	.002	.007	.007	.007	.007	.008	.012	.015

Circular

B

	$BF_{inclusion}$								
TE (A → B)	1.28	🟧	🟦	🟧	🟧	🟧	🟧	🟧	🟧
Open question A	4.67	🟧	🟧	🟧	🟧	🟧	🟧	🟧	🟧
Rating Q1 A	0.55	🟦	🟦	🟦	🟦	🟧	🟦	🟦	🟦
Rating Q2 A	0.57	🟦	🟦	🟦	🟦	🟦	🟧	🟦	🟦
TE (B → A)	0.69	🟦	🟦	🟧	🟧	🟦	🟦	🟧	🟧
Open question B	0.68	🟦	🟦	🟦	🟦	🟦	🟦	🟧	🟦
Rating Q1 B	0.71	🟦	🟦	🟦	🟦	🟦	🟦	🟦	🟧
Rating Q2 B	0.94	🟦	🟦	🟧	🟦	🟦	🟦	🟧	🟦
	BF_{10}	3.83	2.80	2.76	2.59	1.88	1.71	1.63	1.48
	R^2	0.32	0.21	0.42	0.36	0.33	0.32	0.44	0.37
	p	.017	.023	.025	.028	.041	.047	.047	.053

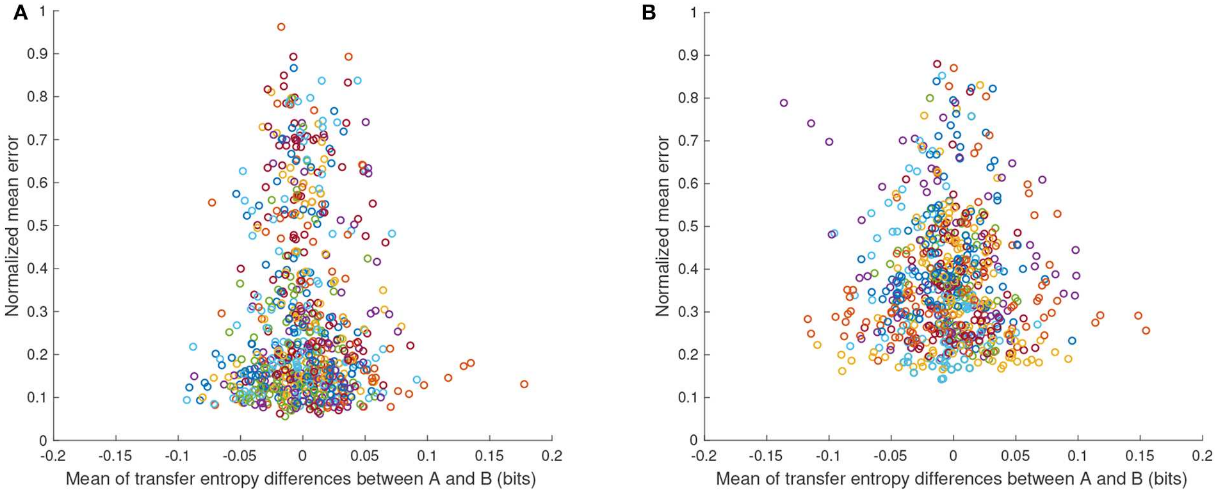
Elliptical

FIGURE 8 | Bayesian linear regression analyses with performance error as criterion variable and *TE*(*A* → *B*), *TE*(*B* → *A*) and subjective measures (open question, Q1 and Q2) from participants A and B as predictor variables. The Bayes factor BF10 describes how likely the data occur under a regression model with the highlighted predictors of one row; the inclusion Bayes factor (BFinclusion) describes how likely the data occur under models that include the respective predictor. **(A)** Circular. **(B)** Elliptical.

higher coordination levels improve task performance, however it is not indispensable. Different, e.g., loosely coupled, social behavior might achieve a good performance as well. The effect of exchanging seats in the elliptic study introduced higher variability in the transfer entropy measure, although the main trends remained consistent.

In summary, the results of both experiments were consistent regarding the validation of TE as objective measure for interpersonal coordination in this task, as well as the significant correlations between TE and joint task performance, and with respect to the identification of leader-follower roles on a descriptive level. Similarly to Takamizawa and Kawasaki (2019), we hypothesized that TE from leader- to follower-like behavior was large and vice versa, from follower to leader—small. When the TE in both directions was on the same level—equally high or equally low—we hypothesized that there was no pronounced leader-follower relationship according to our measure. Different were the correlations between subjective and objective measures. Despite these differences, the results of both studies suggest that TE might be a useful tool for studying factors for subjective experiences in social interactions.

The seat exchange provided one important insight in this study, namely the seat-invariance of social roles within certain dyads, since regardless of the seat the flow of entropy kept the direction from one player to the other (**Figure 9**). Furthermore, the results of the Bayesian linear regression analyses (**Figure 8**), which allow to compare the predictive value of $TE(A \rightarrow B)$ and $TE(B \rightarrow A)$ on the joint task performance, indicate that social roles also depend on the first seating arrangement and preserve when the participants change their seats. This



Circular

Elliptical

FIGURE 9 | Performance (normalized mean absolute error) vs. transfer entropy differences per trial for all pairs from experiment A (**A**, circular target) and from experiment B (**B**, elliptical target). In both studies, the point densities have a characteristic bell shape, highlighting particular trade-offs.

suggests the potential of the applied measure to infer simple social relationships based solely on behavioral data recorded in a smooth rhythmic repetitive interaction. It is well-known that information-theoretic functionals, such as transfer entropy, require large amounts of data in order to provide reliable estimates. Since trials were considerably short, the analysis seem to have been enhanced by the simplicity of interaction.

On the other hand, the simplicity of the experimental paradigm seems to have raised challenges in the subjective evaluation, particularly when addressing the sense of collaboration and interpersonal coordination. We used an explicit open-ended question for establishing the potential leadership within the pairs, as well as a few more subtle indirect questions rated on a Likert scale. The subjective data did not provide consistent trends, which suggests how difficult it was to subjectively evaluate one's own performance in the social aspect of such collaborative interaction. Another issue complicating the subjective data analysis was the fact that questions were answered only once at the end of the experiment and therefore the ratings did not usually apply for all trials (particularly not for the earlier ones). That may have been one potential reason for the participants' confusion in the evaluation, as it was left up to them to decide how to rate the whole experiment (including earlier and later trials alike), providing the otherwise complicated nature of this judgement. In order to make sense of the ratings, we applied a simple (non-distorting) linear transformations on the subjective data which kept the major trends, and correlated the results to the corresponding major trends in the objective measure of coordination (i.e., transfer entropy).

The results of this study suggest the potential of model-free measures of information transfer, such as the transfer entropy, for the analysis of the social aspects of highly interactive collaborative studies, particularly involving simplistic rhythmic controls. Other methods, such as lagged cross-correlation or Granger causality (Granger, 1969), require stationarity and normality assumptions or pre-defined models. Transfer entropy has seen a dramatic surge of interest in neuroscience recently (Wibral et al., 2014), where it is used to estimate the information transfer between two tightly coupled processes. We extend its field of application to less tightly coupled stochastic processes, which form a closed loop with hundreds of milliseconds of lag and involve the full human sensorimotor and decision making hierarchy of control.

The task required the participants to anticipate the combined effect of their joint actions. This might enhance participants' understanding of their own and their partner's actions as well as joint actions, which positively affects interpersonal coordination. In addition, previously published literature (Schmidt and Richardson, 2008; Keller et al., 2014; Lang et al., 2016; Loehr and Vesper, 2016) highlights the significance of rhythmical movement components in interpersonal coordination. Additionally, there is evidence that the rhythmic component during interpersonal coordination reduces practice effort and errors (Lang et al., 2016; Loehr and Vesper, 2016).

Overall, the results supported the hypothesis that this type of collaborative interaction intrinsically motivates the emergence of interpersonal rhythmic coordination, which could be objectively characterized with information-theoretic tools. Our analysis provides quantitative evidence for the emergence of leader-follower relationships, guided by the general principle of perceptual–motor coordination, and is an initial step toward defining more general social sensorimotor contingencies. This evidence was consistent across the experiments and was not diminished by task difficulty levels or seating arrangements of the participants. Similar leader-follower relationships have been identified in other studies (Konvalinka et al., 2014; Takamizawa and Kawasaki, 2019) using different social interaction tasks and/or different analytical methods. Takamizawa and Kawasaki (2019) apply transfer entropy on behavioral data as well, however their study is based on a discrete tapping task, while our study explores a highly interactive continuous paradigm. Our analysis suggests that although subjects tend to steadily improve their coordination skills over time and produce tightly coupled rhythmic control signals, they do not necessarily apply such techniques in order to improve performance.

Finally, the current study has important implications for future research on the social psychological aspects of interpersonal coordination as it reveals the potential of non-parametric information-theoretic methods for quantifying behavioral trends in joint cognitive systems, which are typically identified qualitatively by human observers. The tetherball paradigm provided an easy to learn natural test environment and a basis for examining the interpersonal processes involved in mutual entrainment.

7. CONCLUSION

The characterization of causal dependence can be approached with a variety of methods, and depending on the experimental paradigm in some scenarios certain methods might be more appropriate than others. In this study, we applied the information-theoretic measure of transfer entropy for quantifying the emergent social sensorimotor contingencies in the scope of two studies. While the results look promising, further work is required to explore the range of applicability of this approach for measuring interpersonal coordination in a variety of diverse tasks. Future studies need to address carefully the subjective aspect in the design process. It might be interesting to investigate paradigms in which task performance is inherently correlated with interpersonal coordination. An important aspect for future research is how motor learning and the emergence of interpersonal coordination are related to each other and how that can be expressed with objectives metrics. These aspects are closely related to the perception of kinematics—human control movements or reference object's movements (e.g., a table). Having an objective tool for inferring the level of causal dependence from behavioral data could improve current studies and could facilitate the identification of socializing sensorimotor contingencies in future research.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the Leibniz University Hannover. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

DT wrote the background and the main parts related to the information-theoretic analysis. GS wrote the experimental methodology part. T-HH together with AE and GS developed the paradigm and the experimental design. T-HH realized the software development. AE and GS supervised the data collection. The information-theoretic analysis and major part of the results were realized by DT and DP, supported by GS and T-HH. All authors critically revised the manuscript.

FUNDING

The authors acknowledge support by EU HORIZON2020-FETPROACT-2014 No. 641321.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2020.00185/full#supplementary-material

REFERENCES

Adlakha, S., Lall, S., and Goldsmith, A. (2012). Networked markov decision processes with delays. *IEEE Trans. Autom. Control* 57, 1013–1018. doi: 10.1109/TAC.2011.2168111

Arnold, A., Liu, Y., and Abe, N. (2007). "Temporal causal modelling with graphical granger methods," in *KDD'07* (San Jose, CA), 66. doi: 10.1145/1281192.1281203

Astolfi, L., Toppi, J., Borghini, G., Vecchiato, G., He, E. J., Roy, A., et al. (2012). "Cortical activity and functional hyperconnectivity by simultaneous EEG recordings from interacting couples of professional pilots," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (San Diego, CA), 4752–4755. doi: 10.1109/EMBC.2012.6347029

Astolfi, L., Toppi, J., Borghini, G., Vecchiato, G., Isabella, R., De Vico Fallani, F., et al. (2011). "Study of the functional hyperconnectivity between couples of pilots during flight simulation: an EEG hyperscanning study," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Boston, MA), 2338–2341. doi: 10.1109/IEMBS.2011.6090654

Barnett, L., Barrett, A. B., and Seth, A. K. (2009). Granger causality and transfer entropy are equivalent for gaussian variables. *arXiv* 0910.4514. doi: 10.1103/PhysRevLett.103.238701

Bertschinger, N., Olbrich, E., Ay, N., and Jost, J. (2008). Autonomy: an information-theoretic perspective. *Biosystems* 91, 331–345. doi: 10.1016/j.biosystems.2007.05.018

Bottani, S. (1996). Synchronization of integrate and fire oscillators with global coupling. *Phys. Rev. E* 54, 2334–2350. doi: 10.1103/PhysRevE.54.2334

Box, T. E. P., and Jenkins, G. M. (1970). *Time Series Analysis; Forecasting and Control*. San Francisco, CA: Holden-Day.

Buck, J. (1988). Synchronous rhythmic flashing of fireflies. *Quart. Rev. Biol.* 63:265. doi: 10.1086/415929

Ciaramidaro, A., Toppi, J., Casper, C., Freitag, C., Siniatchkin, M., and Astolfi, L. (2018). Multiple-brain connectivity during third party punishment: an EEG hyperscanning study. *Sci. Rep.* 8:6822. doi: 10.1038/s41598-018-24416-w

Czeszumski, A., Eustergerling, S., Lang, A., Menrath, D., Gerstenberger, M., Schuberth, S., et al. (2020). Hyperscanning: a valid method to study neural inter-brain underpinnings of social interaction. *Front. Hum. Neurosci.* 14:39. doi: 10.3389/fnhum.2020.00039

Demos, A. P., Chaffin, R., Begosh, K. T., Daniels, J. R., and Marsh, K. L. (2012). Rocking to the beat: effects of music and partner's movements on spontaneous interpersonal coordination. *J. Exp. Psychol.Gen.* 141, 49–53. doi: 10.1037/a0023843

Dotov, D., and Froese, T. (2018). Entraining chaotic dynamics: A novel movement sonification paradigm could promote generalization. *Hum. Mov. Sci.* 61, 27–41. doi: 10.1016/j.humov.2018.06.016

Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., and Garnero, L. (2010). Inter-brain synchronization during social interaction. *PLoS ONE* 5:e12166. doi: 10.1371/journal.pone.0012166

Fallani, F., Nicosia, V., Sinatra, R., Astolfi, L., Cincotti, F., Mattia, D., et al. (2010). Defecting or not defecting: How to "read" human behavior during cooperative games by EEG measurements. *PLoS ONE* 5:e14187. doi: 10.1371/journal.pone.0014187

Friston, K., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. *Neuroimage* 19, 1273–1302. doi: 10.1016/S1053-8119(03)00202-7

Fusaroli, R., Raczaszek-Leonardi, J., and Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas Psychol.* 32, 147–157. doi: 10.1016/j.newideapsych.2013.03.005

Fusaroli, R., and Tylén, K. (2015). Investigating conversational dynamics: interactive alignment, interpersonal synergy, and collective task performance. *Cogn. Sci.* 40, 145–171. doi: 10.1111/cogs.12251

Garrod, S., and Pickering, M. J. (2004). Why is conversation so easy? *Trends Cogn. Sci.* 8, 8–11. doi: 10.1016/j.tics.2003.10.016

Garrod, S., and Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Top. Cogn. Sci.* 1, 292–304. doi: 10.1111/j.1756-8765.2009.01020.x

Granger, C. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438. doi: 10.2307/1912791

Haken, H., Kelso, J., and Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biol. Cybern.* 51, 347–356. doi: 10.1007/BF00336922

Harder, M., Polani, D., and Nehaniv, C. L. (2010). "Two agents acting as one," in *Proc. Artificial Life* (Odense).

Hwang, T.-H., Schmitz, G., Klemmt, K., Brinkop, L., Ghai, S., Stoica, M., et al. (2018). Effect- and performance-based auditory feedback on interpersonal coordination. *Front. Psychol.* 9:404. doi: 10.3389/fpsyg.2018.00404

Keller, P. E. (2012). Mental imagery in music performance: underlying mechanisms and potential benefits. *Ann. N. Y. Acad. Sci.* 1252, 206–213. doi: 10.1111/j.1749-6632.2011.06439.x

Keller, P. E., and Appel, M. (2010). Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles. *Music Percept.* 28, 27–46. doi: 10.1525/mp.2010.28.1.27

Keller, P. E., Novembre, G., and Hove, M. J. (2014). Rhythm in joint action: psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20130394. doi: 10.1098/rstb.2013.0394

Kelso, J., Delcolle, J., and Schöner, G. (1990). "Action-perception as a pattern formation process," in *Attention and Performance XIII*, Vol. 5, ed M. Jeannerod (Hillsdale, NJ: Erlbaum), 139–169. doi: 10.4324/9780203772010-5

Kelso, J. S. (1995). *Dynamic Patterns: The Self-Organization of Brain and Behavior* (MIT Press).

Knoblich, G., and Jordan, J. S. (2003). Action coordination in groups and individuals: learning anticipatory control. *J. Exp. Psychol. Learn. Mem. Cogn.* 29, 1006–1016. doi: 10.1037/0278-7393.29.5.1006

Kojima, H., Froese, T., Oka, M., Iizuka, H., and Ikegami, T. (2017). A sensorimotor signature of the transition to conscious social perception: co-regulation of active and passive touch. *Front. Psychol.* 8:1778. doi: 10.3389/fpsyg.2017.01778

Konvalinka, I., Bauer, M., Stahlhut, C., Hansen, L., Roepstorff, A., and Frith, C. (2014). Frontal alpha oscillations distinguish leaders from followers: Multivariate decoding of mutually interacting brains. *Neuroimage* 94, 79–88. doi: 10.1016/j.neuroimage.2014.03.003

Kraskov, A., Stögbauer, H., and Grassberger, P. (2004). Estimating mutual information. *Phys. Rev. E* 69:066138. doi: 10.1103/PhysRevE.69.066138

Kuramoto, Y. (1984). *Chemical Oscillations, Waves, and Turbulences*. Berlin: Springer-Verlag. doi: 10.1007/978-3-642-69689-3

Lang, M., Shaw, D. J., Reddish, P., Wallot, S., Mitkidis, P., and Xygalatas, D. (2016). Lost in the rhythm: effects of rhythmon subsequent interpersonal coordination. *Cogn. Sci.* 40, 1797–1815. doi: 10.1111/cogs.12302

Leong, V., Byrne, E., Clackson, K., Georgieva, S., Lam, S., and Wass, S. (2017). *Speaker Gaze Increases Information Coupling Between Infant and Adult Brains*.

Liu, D., Liu, S., Liu, X., Li, A., Jin, C., and Chen, Y. (2018). Interactive brain activity: review and progress on EEG-based hyperscanning in social interactions. *Front. Psychol.* 9:1862. doi: 10.3389/fpsyg.2018.01862

Liu, N., Mok, C., Witt, E., Pradhan, A., Chen, J., and Reiss, A. (2016). Nirs-based hyperscanning reveals inter-brain neural synchronization during cooperative jenga game with face-to-face communication. *Front. Hum. Neurosci.* 10:82. doi: 10.3389/fnhum.2016.00082

Lizier, J. T. (2014). Jidt: An information-theoretic toolkit for studying the dynamics of complex systems. *Front. Robot. AI* 1:11. doi: 10.3389/frobt.2014.00011

Lobier, M., Siebenhühner, F., Palva, S., and Palva, M. (2013). Phase transfer entropy: a novel phase-based measure for directed connectivity in networks coupled by oscillatory interactions. *Neuroimage* 85, 853–872. doi: 10.1016/j.neuroimage.2013.08.056

Loehr, J. D., and Vesper, C. (2016). The sound of you and me: novices represent shared goals in joint action. *Q. J. Exp. Psychol.* 69, 535–547. doi: 10.1080/17470218.2015.1061029

Marsh, K. L., Richardson, M. J., and Schmidt, R. C. (2009). Social connection through joint action and interpersonal coordination. *Top. Cogn. Sci.* 1, 320–339. doi: 10.1111/j.1756-8765.2009.01022.x

Massey, J. L. (1990). "Causality, feedback and directed information," in *Proceedings of the International Symposium on Information Theory and its Applications* (Waikiki, HI).

Naeem, M., Mcginnity, T., Watson, D., Wong-Lin, K., Prasad, G., and Kelso, S. (2012). *Inter-Brain Mutual Information in Social Interaction Tasks* (London: IEEE), 25–28. doi: 10.1109/PRNI.2012.32

Nalatore, H., Ding, M., and Rangarajan, G. (2007). Mitigating the effects of measurement noise on granger causality. *Phys. Rev. E* 75:031123. doi: 10.1103/PhysRevE.75.031123

Newtson, D., Hairfield, J., Bloomingdale, J., and Cutino, S. (1987). The structure of action and interaction. *Soc. Cogn.* 5, 191–237. doi: 10.1521/soco.1987.5.3.191

Pan, Y., Cheng, X., Zhang, Z., Li, X., and Hu, Y. (2017). Cooperation in lovers: an fNIRS-based hyperscanning study. *Hum. Brain Mapp.* 38, 831–841. doi: 10.1002/hbm.23421

Peskin, C. S. (1975). *Mathematical Aspects of Heart Physiology*. New York, NY: Courant Institute of Mathematical Sciences.

Phillips-Silver, J., and Keller, P. (2012). Searching for roots of entrainment and joint action in early musical interactions. *Front. Hum. Neurosci.* 6:26. doi: 10.3389/fnhum.2012.00026

Pikovsky, A., Rosenblum, M., and Kurths, J. (2001). *Synchronization: A Universal Concept in Nonlinear Science*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511755743

Redcay, E., Dodell-Feder, D., Pearrow, M., Mavros, P., and Kleiner, M. (2010). Live face-to-face interaction during fMRI: a new tool for social cognitive neuroscience. *Neuroimage* 50, 1639–1647. doi: 10.1016/j.neuroimage.2010.01.052

Richardson, M., Marsh, K., and Schmidt, R. (2005). Effects of visual and verbal interaction on unintentional interpersonal coordination. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 62–79. doi: 10.1037/0096-1523.31.1.62

Ryali, S., Supekar, K., Chen, T., and Menon, V. (2011). Multivariate dynamical systems models for estimating causal interactions in fmri. *Neuroimage* 54, 807–823. doi: 10.1016/j.neuroimage.2010.09.052

Sänger, J., Müller, V., and Lindenberger, U. (2012). Intra-and interbrain synchronization and network properties when playing guitar in duets. *Front. Hum. Neurosci.* 6:312. doi: 10.3389/fnhum.2012.00312

Sänger, J., Müller, V., and Lindenberger, U. (2013). Directionality in hyperbrain networks discriminates between leaders and followers in guitar duets. *Front. Hum. Neurosci.* 7:234. doi: 10.3389/fnhum.2013.00234

Schippers, M., Roebroeck, A., Renken, R., Nanetti, L., and Keysers, C. (2010). Mapping the information flow from one brain to another during gestural communication. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9388–9393. doi: 10.1073/pnas.1001791107

Schmidt, R., Bienvenu, M., Fitzpatrick, P., and Amazeen, P. (1998). A comparison of intra- and interpersonal interlimb coordination: coordination breakdowns and coupling strength. *J. Exp. Psychol. Hum. Percept. Perform*. 24, 884–900. doi: 10.1037/0096-1523.24.3.884

Schmidt, R. C., Carello, C., and Turvey, M. T. (1990). Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. *J. Exp. Psychol. Hum. Percept. Perform*. 16, 227–247. doi: 10.1037/0096-1523.16.2.227

Schmidt, R. C., and O'Brien, B. (1997). Evaluating the dynamics of unintended interpersonal coordination. *Ecol. Psychol*. 9, 189–206. doi: 10.1207/s15326969eco0903_2

Schmidt, R. C., and Richardson, M. J. (2008). "Dynamics of interpersonal coordination," in *Understanding Complex Systems* (Berlin; Heidelberg: Springer), 281–308. doi: 10.1007/978-3-540-74479-5_14

Schmidt, R. C., and Turvey, M. T. (1994). Phase-entrainment dynamics of visually coupled rhythmic movements. *Biol. Cybern*. 70, 369–376. doi: 10.1007/BF00200334

Schreiber, T. (2000). Measuring information transfer. *Phys. Rev. Lett*. 85, 461–464. doi: 10.1103/PhysRevLett.85.461

Sebanz, N., Bekkering, H., and Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends Cogn. Sci*. 10, 70–76. doi: 10.1016/j.tics.2005.12.009

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J*. 27, 379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x

Stephens, R., and Galloway, T. (2017). Are neurodynamic organizations a fundamental property of teamwork? *Front. Psychol*. 8:644. doi: 10.3389/fpsyg.2017.00644

Strogatz, S. (2003). *Sync: The Emerging Science of Spontaneous Order*. New York, NY: Hyperion Press.

Takamizawa, K., and Kawasaki, M. (2019). Transfer entropy for synchronized behavior estimation of interpersonal relationships in human communication: identifying leaders or followers. *Sci. Rep*. 9:10960. doi: 10.1038/s41598-019-47525-6

Valdes-Sosa, P., Sanchez-Bornot, J., Lage-Castellanos, A., Vega-Hernandez, M., Bosch-Bayard, J., Melie-Garcia, L., et al. (2005). Estimating brain functional connectivity with sparse multivariate autoregression. *Philos. Trans. R. Soc. Lond. B Biol. Sci*. 360, 969–981. doi: 10.1098/rstb.2005.1654

Vicaria, I., and Dickens, L. (2016). Meta-analyses of the intra-and interpersonal outcomes of interpersonal coordination. *J. Nonverbal Behav*. 40, 335–361. doi: 10.1007/s10919-016-0238-8

Walker, T. J. (1969). Acoustic synchrony: two mechanisms in the snowy tree cricket. *Science* 166:891. doi: 10.1126/science.166.3907.891

Waterhouse, E., Watts, R., and Blăsing, B. E. (2014). Doing duo-a case study of entrainment inwilliam forsythe's choreography "duo". *Front. Hum. Neurosci*. 8:812. doi: 10.3389/fnhum.2014.00812

Wibral, M., Vicente, R., and Lindner, M. (2014). *Transfer Entropy in Neuroscience*. Berlin; Heidelberg: Springer Berlin Heidelberg, 3–36. doi: 10.1007/978-3-642-54474-3_1

Winfree, A. (1967). Biological rhythms and behavior of populations of coupled oscillators. *J. Theor. Biol*. 16, 15–42. doi: 10.1016/0022-5193(67)90051-3

Winfree, A. (1980). *The Geometry of Biological Time*. New York, NY: Springer-Verlag. doi: 10.1007/978-3-662-22492-2

Yun, K., Watanabe, K., and Shimojo, S. (2012). Interpersonal body and neural synchronization as a marker of implicit social interaction. *Sci. Rep*. 2:959. doi: 10.1038/srep00959

Ziebart, B. D. (2013). "Robust structure estimation of maximum causal entropy processes," in *51st Annual Allerton Conference on Communication, Control, and Computing* (Monticello, IL). doi: 10.1109/Allerton.2013.6736634

Ziebart, B. D., Bagnell, J. A., and Dey, A. K. (2013). The principle of maximum causal entropy for estimating interacting processes. *IEEE Trans. Inform. Theory* 59, 1966–1980. doi: 10.1109/TIT.2012.2234824

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Human Neuroscience

PERSPECTIVE
published: 26 May 2020
doi: 10.3389/fnhum.2020.00168

Informing, Coordinating, and Performing: A Perspective on Functions of Sensorimotor Communication

Cordula Vesper[1,2]* *and Vassilis Sevdalis*[3]

[1]Department of Linguistics, Cognitive Science and Semiotics, Aarhus University, Aarhus, Denmark, [2]Interacting Minds Centre, Aarhus University, Aarhus, Denmark, [3]Department of Public Health, Sport Science, Aarhus University, Aarhus, Denmark

Sensorimotor communication is a form of communication instantiated through body movements that are guided by both instrumental, goal-directed intentions and communicative, social intentions. Depending on the social interaction context, sensorimotor communication can serve different functions. This article aims to disentangle three of these functions: (a) an informing function of body movements, to highlight action intentions for an observer; (b) a coordinating function of body movements, to facilitate real-time action prediction in joint action; and (c) a performing function of body movements, to elicit emotional or aesthetic experiences in an audience. We provide examples of research addressing these different functions as well as some influencing factors, relating to individual differences, task characteristics, and situational demands. The article concludes by discussing the benefits of a closer dialog between separate lines of research on sensorimotor communication across different social contexts.

Keywords: sensorimotor communication, joint action, nonverbal communication, action prediction, dance, music, sport, aesthetics

Edited by:
Alfred Oliver Effenberg,
Leibniz University Hannover,
Germany

Reviewed by:
Nathalie George,
Centre National de la Recherche
Scientifique (CNRS), France
Jennifer MacRitchie,
Western Sydney University, Australia

****Correspondence:***
Cordula Vesper
cvesper@cc.au.dk

Specialty section:
This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience

Received: *25 January 2020*
Accepted: *20 April 2020*
Published: *26 May 2020*

Citation:
Vesper C and Sevdalis V (2020) Informing, Coordinating, and Performing: A Perspective on Functions of Sensorimotor Communication.
Front. Hum. Neurosci. 14:168.
doi: 10.3389/fnhum.2020.00168

INTRODUCTION

Humans have an intrinsic ability to interact socially with others. From an early age, and before cultivating the language faculty, humans are able to understand others and to be understood by others through pre- and nonverbal cues, such as pointing gestures and gaze direction (Tomasello, 2019). Later in life, this ability becomes particularly relevant in social contexts where the environment prevents verbal exchange (e.g., due to background noise) or where linguistic forms of communication are not appropriate (e.g., in sport and performing art contexts). In such cases, individuals express meaning through their actions and body movements. Nonverbal forms of communication are pertinent in human cultures worldwide (Matsumoto, 2006) and occur in various contexts, from complementing or replacing verbal communication in everyday interactions (Vesper and Richardson, 2014; Peeters et al., 2015; Vesper et al., 2017b; Pezzulo et al., 2019) to supporting complex interpersonal interactions and producing art through dance and music (Sevdalis and Keller, 2011a, 2014; D'Ausilio et al., 2015; MacRitchie et al., 2017; Bishop et al., 2019).

In the past decades, considerable attention has been placed on understanding the foundations of cognitive and social processes within human actions and embodied interactions (Gallese, 2007).

Within an embodied cognition framework, bodily movements and sensorimotor experiences are considered pivotal in shaping cognitive functions such as learning, memory, and perception (Wilson, 2002; Barsalou, 2008). One consequence of the embodied nature of cognition is that individuals employ their sensorimotor skills when observing the actions of other individuals (Blakemore and Frith, 2005; Wilson and Knoblich, 2005; Grafton, 2009; Schubert and Semin, 2009). This action simulation or motor resonance is regarded as a fundamental mechanism for social cognition, bridging the gap between self and other (Prinz, 1990; Jeannerod, 2006; Vesper et al., 2010; Herwig et al., 2013). Specifically, the direct matching between action and perception can act as a foundation for the coupling of individual minds and the emergence of sensorimotor communication between them.

In contrast to many predominantly communicative actions such as gesturing while speaking, sensorimotor communication is instantiated through actions that are guided by both communicative, social intentions and by instrumental, goal-directed intentions. To illustrate how this double nature of sensorimotor communication can serve different functions, consider the following example: While continuously playing her musical instrument and producing a desired complex sound pattern (an instrumental action goal), an ensemble musician can inform another performer about her intention to enter a specific musical passage by exaggerating the movement of her upper body (a communicative action goal). A second musician can understand this intention and respond by slowing down the musical tempo so that they play together in synchrony. For an observing audience, the musicians' coordinated movements can elicit aesthetic experiences and emotional reactions. The musicians' expertise, their experience with each other's playing style, and their shared musical and cultural backgrounds can all influence their resulting performance (cf. Keller, 2014).

A multitude of research studies identified kinematic parameters such as movement amplitude or grasp size that are modified depending on an agent's action intention (for an overview and discussion, see Ansuini et al., 2014). Previous research on sensorimotor communication examined how such kinematic parameters are modified in joint action to be informative for a co-actor. To that end, a computational model linked movement modifications to internal predictive processes and postulated that sensorimotor communication serves the purpose of facilitating prediction for an observer (Pezzulo et al., 2013). Moreover, a recent framework classified various forms of verbal and nonverbal information exchange (Pezzulo et al., 2019). With this article, we intend to complement such accounts by focusing on the functions that sensorimotor communication serves within different social interaction contexts. In particular, we distinguish three central functions—informing, coordinating, and performing—that differ in the directionality of information flow between individuals, as illustrated in **Figure 1**.

Accordingly, the section "Informing Function: Highlighting Own Action Intentions for an Observer" introduces movements that provide relevant information to another person, by highlighting an individual's action intentions. The section "Coordinating Function: Facilitating Real-Time Action Prediction in Joint Action" continues with movements that allow close coordination between multiple individuals' actions, by facilitating mutual predictions in real-time. The section "Performing Function: Eliciting Emotional or Aesthetic Experiences in an Audience" addresses movements that support performance in contexts such as music and dance, by conveying dynamic expressive nuances that elicit emotional and aesthetical experiences in an audience. These three functions of sensorimotor communication can be influenced by several factors related to individual differences (e.g., in self-report measures of empathy or sensorimotor expertise), specific task characteristics (e.g., when interacting individuals have

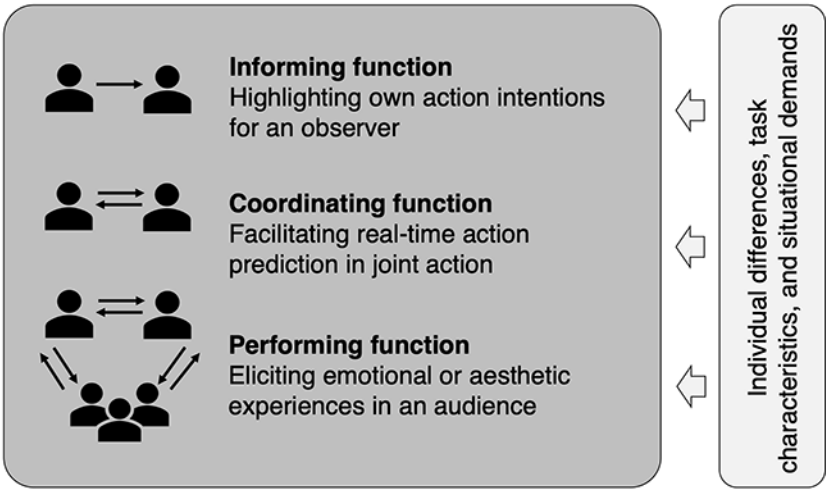


FIGURE 1 | An illustration of three central functions of sensorimotor communication, indicating the main directionality of information exchange between individuals and potential influencing factors on their communication. The number of involved individuals can vary in each context.

asymmetric access to task-relevant information), or situational demands (e.g., in a music performance context). To conclude, the section "Final Remarks" discusses the potential benefits of a closer dialog between separate lines of research addressing sensorimotor communication.

INFORMING FUNCTION: HIGHLIGHTING OWN ACTION INTENTIONS FOR AN OBSERVER

The first function of sensorimotor communication, that we address in this article, is providing another person with information about one's intended body movements. It is well established that individual observers are sensitive to information about others' movement intentions, allowing them to predict what another person will do next (Graf et al., 2007; Becchio et al., 2012; Cavallo et al., 2016). For example, observers can reliably distinguish different social intentions towards another person, such as giving instructions or requesting information (Manera et al., 2010). Factors such as movement complexity, amount of visual information about the movement, and exposure duration to the movement have frequently been related to higher recognition rates (Pollick et al., 2003; Dahl and Friberg, 2007; Sevdalis and Keller, 2009, 2010).

Given this evidence, it could be argued that movements are informative *per se*. In many social contexts, however, movement information is not just passively transmitted as a byproduct of acting; instead, individuals often deliberately modify their movements to make their action intention visible to others. Thus, movements are intentionally modified to be (even more) informative. This function of sensorimotor communication occurs most prominently in situations where making another person aware of one's intention is explicitly desired, such as during teaching and demonstration. For instance, a dance teacher might exaggerate movement cues to make her pupils understand what is most important to imitate, while she continues to perform the dance movement itself. Empirical research reveals the flexibility with which certain features of an action can be modified to fulfill the informative function of sensorimotor communication. As an example, individuals, who teach an observer a particular musical sequence, exaggerate kinematic features such as amplitude and velocity so that they become informative about where in space the movement is directed towards (McEllin et al., 2018). Similarly, in child-directed action, it has been shown that adults tend to modify movement cues to teach a child, for example, how to use a novel tool ("motionese"; Brand et al., 2002). This sensorimotor communication is thought to support learning by highlighting and separating the relevant action steps (Koterba and Iverson, 2009; Williamson and Brand, 2014). Even minimal modifications are sufficient for supporting the recognition of intentions. For instance, movements produced in a joint action context often contain sufficient information, so that even individuals unrelated to a specific interaction context can predict the actors' movement goals from simple static images (Vesper and Richardson, 2014) or temporal cues (Vesper et al., 2017b).

Although most research has investigated communicative action modulations in cooperative contexts, they also occur in competitive contexts, where movements are intentionally modified to be less informative or misleading. One domain is competitive sports, where players might encounter deceptive body movements from their opponents. By deliberately providing "fake" information about one's action intention, players can attempt to disturb an opponent's prediction processes (Cañal-Bruland, 2017), misleading them about the upcoming action and, therefore, eliciting an inappropriate response. In cases such as handball, rugby, and basketball, expert performers have demonstrated a perceptual advantage in correctly disambiguating others' movement intentions (Cañal-Bruland and Schmidt, 2009; Sebanz and Shiffrar, 2009; Brault et al., 2012; Mori and Shimada, 2013), highlighting the role of action expertise in sensorimotor communication.

COORDINATING FUNCTION: FACILITATING REAL-TIME ACTION PREDICTION IN JOINT ACTION

Beyond merely providing information, body movements can also support real-time coordination between multiple individuals' actions. In joint action settings, it is often not only necessary to understand the partner's immediate action intention, that is, what this person is going to do next, but also to be able to perform an appropriate complementary action at the right time, that is, to choose which respective action to perform and when to act. In other words, the real-time constraints and mutual influences between co-actors place a high burden on joint planning and performance. Sensorimotor communication, here, plays the role of a "coordination smoother" (Vesper et al., 2010), a way of simplifying coordination.

One of the most studied contexts of this coordinative function of sensorimotor communication is that of achieving synchrony between two individuals' actions. In a study with expert pianists playing duets, it was observed that restricting access to shared auditory information made the players visually enhance their finger movement height, which, in turn, allowed them to maintain precise temporal coordination with each other (Goebl and Palmer, 2009). Sensorimotor communication, in this case, compensated for missing auditory information through another (here: visual) modality. Similar findings were obtained in a joint sequence coordination task, where those persons in a dyad who received prior information about upcoming target locations, deliberately modulated their movement amplitude while interacting (Vesper and Richardson, 2014). In particular, they moved to relatively far targets with a higher amplitude and a different velocity profile than to relatively close locations. This modulation allowed the co-actors, who did not receive prior information, to anticipate the location of the correct target and their partners' movements more efficiently. The impact of sensorimotor communication on the outcome of a joint action was directly tested in a study that modulated the type of perceptual information shared between co-actors (Vesper et al., 2016). Pairs of participants synchronized the endpoints of simple target-directed movements. Compared to a condition without visual access to each other's movements, participants exaggerated the amplitude of their movements in a condition with visual access. This in turn, allowed them to be more synchronized. Other action features besides movement amplitude are modulated for communication, including grasp position on an object (Schmitz et al., 2018), grasp aperture while moving towards an object (Sacheli et al., 2013), and the speed with which to approach a target location (Vesper et al., 2017b).

Individual differences in co-actors' social skills or their interaction roles can influence the execution of body movements. Several studies demonstrated that leader/follower relationships influence how dyads or larger groups approach a task, and how well they manage to adjust their movements and achieve fine temporal and spatial coordination with each other (Konvalinka et al., 2010; Glowinski et al., 2013; Badino et al., 2014; Curioni et al., 2019). The assignment of leadership roles can also alter the duration of gaze towards the co-performer in duetting pianists (e.g., before tempo changes), affecting how well they achieve musical synchronization (Kawase, 2014). Coordination of music and body movement can also be influenced by familiarity with a co-performer's musical part and previous rehearsals (Williamon and Davidson, 2002; Ragert et al., 2013; Keller, 2014). Similar effects likely extend into movement performance in contexts that foster sensorimotor communication. One study tested the influence of inter-individual differences on the emergence of new communication systems (Volman et al., 2012). Their findings suggest that individuals tend to differ in their ability to understand another person's movement intentions, which, in turn, can affect the success with which dyads manage to creatively invent non-conventional ways of communicating.

PERFORMING FUNCTION: ELICITING EMOTIONAL OR AESTHETIC EXPERIENCES IN AN AUDIENCE

In the interaction contexts described in the previous sections, the focus is on informing or transferring task-relevant information to another individual or completing tasks together. In contrast, many situations, such as dance and music performances, can generate more complex interactions, where performers engage in deliberate modulations of their bodily movements to convey meaning to an audience, such as eliciting emotions, aesthetical experiences, or narratives in an observer's mind. As illustrated in **Figure 1**, this transforms the overall interactive structure by adding more roles (i.e., performer and audience) and, possibly, more people (although this does not exclude contexts with only one performer, more than two performers, or varying sizes of the audience).

Performers' movements are potent carriers of aesthetic significance and often convey spatial and temporal expression dynamics in visual and auditory modalities, which influence the experience of observers (Vines et al., 2006; MacRitchie et al., 2013). For example, musical performances are judged as more interesting when, in addition to hearing the music, observers can see the musicians playing in an expressive compared to an inexpressive manner (Broughton and Stevens, 2009). Similarly, dance movements depicting greater displacement of a dancer's body in space are associated with higher liking ratings from spectators (Calvo-Merino et al., 2008). In contrast to static movement displays, dynamic displays of movements across time provide audiences with information that allow them to infer information about a performer's intended artistic expression which, for example, allows them to differentiate between expressive and inexpressive motion cues in performing musicians or dancers (Sevdalis and Keller, 2012; MacRitchie et al., 2017). Moreover, aesthetic responses to dance movements can be intensified by dance performers' interpersonal synchrony (Vicary et al., 2017). Thus, body movements are effective channels for the communication of performers' intentions for expression, as well as for inducing aesthetic experiences in an observing audience.

Various factors can influence how accurately others' intentions about expression intensity and aesthetic significance can be perceived in performance contexts. One important factor is performers' and observers' sensorimotor expertise, which can emerge from the long-term cultivation of a sensorimotor skill and deliberate practice. In the field of music, for example, expert pianists, organ players, and orchestral conductors were shown to be able to reliably distinguish whether recordings of music performances involve previously executed actions of themselves or other individuals (Keller et al., 2007; Gingras et al., 2011; Wöllner, 2012a). Another important factor, apart from long-term or domain-specific expertise, is incidental sensorimotor experience with an action or an interaction partner, which can be beneficial in a communication process. When observers were asked to identify the intended expression intensity of non-expert dancers, recognition accuracy differed depending on whether the observers had motor experience (observing their own dancing movements), visual experience (observing movements of a dancing partner), or no experience with the displayed actions (observing the dancing movements of a stranger; Sevdalis and Keller, 2011b). A further example of individual characteristics that may influence sensorimotor communication is trait-like individual differences, for example, related to empathy. Individuals scoring higher on empathy in self-report questionnaires were found to be more accurate in estimating performers' intentions for expression, whether they were observing ensemble musicians (Wöllner, 2012b) or dancers (Sevdalis and Keller, 2012; Sevdalis and Raab, 2014; Sevdalis and Raab, 2016).

FINAL REMARKS

The present article illustrates the complexity that research on sensorimotor communication in social interaction needs to address: Studying only simple information exchange may not suffice to fully understand how body movements are used to facilitate coordination between individuals; studying only coordination of body movements in a dyadic setting may not suffice to understand the dynamics of interaction between co-performers' expressions and an audience's aesthetic responses; focusing only on complex applied performance contexts may not benefit from basic research that addresses low-level sensorimotor processes. Thus, to investigate such complexity, future research will benefit from a multidisciplinary dialog across fields as varied as human movement science, joint action, communication studies, and performance psychology (Vesper et al., 2010, 2017a; D'Ausilio et al., 2015; Sevdalis and Wöllner, 2016).

To date, despite the considerable potential of crosstalk between different fields that focus on communicative actions, the systematic investigation of movements still receives less attention compared to the study of other cognitive processes (Rosenbaum and Feghhi, 2019). Accordingly, to complement the literature on basic motor processes in sensorimotor communication (e.g., Pezzulo et al., 2013; Vesper and Richardson, 2014; Vesper et al., 2017b), we aim to extend the discussion to studies illustrating how movements serve as carriers of meaning and expression dynamics in performing arts contexts (Sevdalis and Keller, 2011a, 2014; Sevdalis and Wöllner, 2016). We hope our proposed schema in **Figure 1** will support the systematic assessment of different parameters in human social interaction, such as individuals' goals that necessitate certain body movements, the contexts in which the movements are embedded, or the particular characteristics of the individuals executing and perceiving these movements. In particular, future research could directly compare contexts that differ only in the function that sensorimotor communication has, and, thereby, specify which influencing factors are particularly relevant to which context.

In our view, the study of sensorimotor communication can enhance our understanding of human cognitive processes more generally, by offering an interactive approach, where cognitive processes do not lie just in one individual mind, but where an acting individual/performer and a partner/audience are investigated as participatory agents in a large-scale communication process. As individuals possess considerable abilities in providing social information to others through their movements and in inferring social information from the subtle movement cues of others, this eventual attunement to mutually exchanged cues can be regarded as a fundamental characteristic of the sensorimotor basis of human social cognition. Besides, going beyond human social processes, a systematic investigation of different functions of sensorimotor communication also promises to be informative for applied research on artificial agents. Just as when two or more humans work together in proximity, and with high temporal and spatial precision, humans interacting with robots may also benefit from the direct and fast information exchange instantiated through nonverbal communicative cues (Dragan and Srinivasa, 2014; Vesper, 2014; Donnarumma et al., 2018). The future will tell how similar robot behavior needs to be to human behavior, to allow the same smooth and easy interaction that we see when humans play basketball, perform a Bach cantata, or simply shake hands with each other.

AUTHOR CONTRIBUTIONS

CV and VS contributed equally to the conceptual development and writing of the article.

REFERENCES

Ansuini, C., Cavallo, A., Bertone, C., and Becchio, C. (2014). The visible face of intention: why kinematics matters. *Front. Psychol.* 5:815. doi: 10.3389/fpsyg.2014.00815

Badino, L., D'Ausilio, A., Glowinski, D., Camurri, A., and Fadiga, L. (2014). Sensorimotor communication in professional quartets. *Neuropsychologia* 55, 98–104. doi: 10.1016/j.neuropsychologia.2013.11.012

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645. doi: 10.1146/annurev.psych.59.103006.093639

Becchio, C., Manera, V., Sartori, L., Cavallo, A., and Castiello, U. (2012). Grasping intentions: from thought experiments to empirical evidence. *Front. Hum. Neurosci.* 6:117. doi: 10.3389/fnhum.2012.00117

Bishop, L., Cancino-Chacón, C., and Goebl, W. (2019). Moving to communicate, moving to interact: patterns of body motion in musical duo performance. *Music Percept.* 37, 1–25. doi: 10.1525/mp.2019.37.1.1

Blakemore, S.-J., and Frith, C. (2005). The role of motor contagion in the prediction of action. *Neuropsychologia* 43, 260–267. doi: 10.1016/j.neuropsychologia.2004.11.012

Brand, R. J., Baldwin, D. A., and Ashburn, L. A. (2002). Evidence for 'motionese': modifications in mothers' infant-directed action. *Dev. Sci.* 5, 72–83. doi: 10.1111/1467-7687.00211

Brault, S., Bideau, B., Kulpa, R., and Craig, C. M. (2012). Detecting deception in movement: the case of the side-step in rugby. *PLoS One* 7:e37494. doi: 10.1371/journal.pone.0037494

Broughton, M., and Stevens, C. (2009). Music, movement and marimba: an investigation of the role of movement and gesture in communicating musical expression to an audience. *Psychol. Music* 37, 137–153. doi: 10.1177/0305735608094511

Calvo-Merino, B., Jola, C., Glaser, D. E., and Haggard, P. (2008). Towards a sensorimotor aesthetics of performing art. *Conscious. Cogn.* 17, 911–922. doi: 10.1016/j.concog.2007.11.003

Cañal-Bruland, R. (2017). Deception detection in action: embodied simulation in anti-social human interactions. *Front. Psychol.* 8:166. doi: 10.3389/fpsyg.2017.00166

Cañal-Bruland, R., and Schmidt, M. (2009). Response bias in judging deceptive movements. *Acta Psychol.* 130, 235–240. doi: 10.1016/j.actpsy.2008.12.009

Cavallo, A., Koul, A., Ansuini, C., Capozzi, F., and Becchio, C. (2016). Decoding intentions from movement kinematics. *Sci. Rep.* 6:37036. doi: 10.1038/srep37036

Curioni, A., Vesper, C., Knoblich, G., and Sebanz, N. (2019). Reciprocal information flow and role distribution support joint action coordination. *Cognition* 187, 21–31. doi: 10.1016/j.cognition.2019.02.006

D'Ausilio, A., Novembre, G., Fadiga, L., and Keller, P. E. (2015). What can music tell us about social interaction? *Trends Cogn. Sci.* 19, 111–114. doi: 10.1016/j.tics.2015.01.005

Dahl, S., and Friberg, A. (2007). Visual perception of expressiveness in musicians' body movements. *Music Percept.* 24, 433–454. doi: 10.1525/mp.2007.24.5.433

Donnarumma, F., Dindo, H., and Pezzulo, G. (2018). Sensorimotor communication for humans and robots: improving interactive skills by sending coordination signals. *IEEE Trans. Cogn. Dev. Syst.* 10, 903–917. doi: 10.1109/tcds.2017.2756107

Dragan, A., and Srinivasa, S. (2014). Integrating human observer inferences into robot motion planning. *Auton. Robots* 37, 351–368. doi: 10.1007/s10514-014-9408-x

Gallese, V. (2007). Before and below 'theory of mind': embodied simulation and the neural correlates of social cognition. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 659–669. doi: 10.1098/rstb.2006.2002

Gingras, B., Lagrandeur-Ponce, T., Giordano, B. L., and McAdams, S. (2011). Perceiving musical individuality: performer identification is dependent on performer expertise and expressiveness, but not on listener expertise. *Perception* 40, 1206–1220. doi: 10.1068/p6891

Glowinski, D., Mancini, M., Cowie, R., Camurri, A., Chiorri, C., and Doherty, C. (2013). The movements made by performers in a skilled quartet: a distinctive pattern, and the function that it serves. *Front. Psychol.* 4. doi: 10.3389/fpsyg.2013.00841

Goebl, W., and Palmer, C. (2009). Synchronization of timing and motion among performing musicians. *Music Percept.* 26, 427–438. doi: 10.1525/mp.2009.26.5.427

Graf, M., Reitzner, B., Corves, C., Casile, A., Giese, M., and Prinz, W. (2007). Predicting point-light actions in real-time. *NeuroImage* 36, T22–T32. doi: 10.1016/j.neuroimage.2007.03.017

Grafton, S. T. (2009). Embodied cognition and the simulation of action to understand others. *Ann. N Y Acad. Sci.* 1156, 97–117. doi: 10.1111/j.1749-6632.2009.04425.x

Herwig, A., Beisert, M., and Prinz, W. (2013). "Action science emerging: introduction and leitmotifs," in *Action Science: Foundations of An Emerging Discipline*, eds W. Prinz, M. Beisert and A. Herwig (Cambridge, MA: MIT Press), 1–33.

Jeannerod, M. (2006). *Motor Cognition: What Actions Tell the Self.* New York, NY: Oxford University Press.

Kawase, S. (2014). Assignment of leadership role changes performers' gaze during piano duo performances. *Ecol. Psychol.* 26, 198–215. doi: 10.1080/10407413.2014.929477

Keller, P. E. (2014). "Ensemble performance: interpersonal alignment of musical expression," in *Expressiveness in Music Performance: Empirical Approaches Across Styles and Cultures*, eds D. Fabian, R. Timmers and E. Schubert (Oxford: Oxford University Press), 260–282.

Keller, P. E., Knoblich, G., and Repp, B. H. (2007). Pianists duet better when they play with themselves: on the possible role of action simulation in synchronization. *Conscious. Cogn.* 16, 102–111. doi: 10.1016/j.concog.2005.12.004

Konvalinka, I., Vuust, P., Roepstorff, A., and Frith, C. D. (2010). Follow you, follow me: continuous mutual prediction and adaptation in joint tapping. *Q. J. Exp. Psychol.* 63, 2220–2230. doi: 10.1080/17470218.2010.497843

Koterba, E. A., and Iverson, J. M. (2009). Investigating motionese: the effect of infant-directed action on infants' attention and object exploration. *Infant Behav. Dev.* 32, 437–444. doi: 10.1016/j.infbeh.2009.07.003

MacRitchie, J., Buck, B., and Bailey, N. J. (2013). Inferring musical structure through bodily gestures. *Music. Sci.* 17, 86–108. doi: 10.1177/1029864912467632

MacRitchie, J., Varlet, M., and Keller, P. (2017). "Embodied expression through entrainment and co-representation in musical ensemble performance," in *The Routledge Companion to Embodied Music Interaction*, 1st Edn., eds M. Lesaffre, P.-J. Maes and M. Leman (London: Routledge), 150–159.

Manera, V., Schouten, B., Becchio, C., Bara, B. G., and Verfaillie, K. (2010). Inferring intentions from biological motion: a stimulus set of point light communicative interactions. *Behav. Res. Methods* 42, 168–178. doi: 10.3758/brm.42.1.168

Matsumoto, D. (2006). "Culture and nonverbal behavior," in *Handbook of Nonverbal Communication*, eds V. Manusov and M. Patterson (Thousand Oaks, CA: Sage), 219–235.

McEllin, L., Knoblich, G., and Sebanz, N. (2018). Distinct kinematic markers of demonstration and joint action coordination? Evidence from virtual xylophone playing. *J. Exp. Psychol. Hum. Percept. Perform.* 44, 885–897. doi: 10.1037/xhp0000505

Mori, S., and Shimada, T. (2013). Expert anticipation from deceptive action. *Atten. Percept. Psychophys.* 75, 751–770. doi: 10.3758/s13414-013-0435-z

Peeters, D., Chu, M., Holler, J., Hagoort, P., and Özyurek, A. (2015). Electrophysiological and kinematic correlates of communicative intent in the planning and production of pointing gestures and speech. *J. Cogn. Neurosci.* 27, 2352–2368. doi: 10.1162/jocn_a_00865

Pezzulo, G., Donnarumma, F., and Dindo, H. (2013). Human sensorimotor communication: a theory of signaling in online social interactions. *PLoS One* 8:e79876. doi: 10.1371/journal.pone.0079876

Pezzulo, G., Donnarumma, F., Dindo, H., D'Ausilio, A., Konvalinka, I., and Castelfranchi, C. (2019). The body talks: sensorimotor communication and its brain and kinematic signatures. *Phys. Life Rev.* 28, 1–21. doi: 10.1016/j.plrev.2018.06.014

Pollick, F. E., Hill, H., Calder, A., and Paterson, H. (2003). Recognising facial expression from spatially and temporally modified movements. *Perception* 32, 813–826. doi: 10.1068/p3319

Prinz, W. (1990). "A common coding approach to perception and action," in *Relationships Between Perception and Action: Current Approaches*, eds O. Neumann and W. Prinz (Berlin: Springer), 167–201.

Ragert, M., Schröder, T., and Keller, P. E. (2013). Knowing too little or too much: the effects of familiarity with a co-performer's part on interpersonal coordination in musical ensembles. *Front. Psychol.* 4:368. doi: 10.3389/fpsyg.2013.00368

Rosenbaum, D. A., and Feghhi, I. (2019). The time for action is at hand. *Atten. Percept. Psychophys.* 81, 2123–2138. doi: 10.3758/s13414-018-01647-7

Sacheli, L., Tidoni, E., Pavone, E., Aglioti, S., and Candidi, M. (2013). Kinematics fingerprints of leader and follower role-taking during cooperative joint actions. *Exp. Brain Res.* 226, 473–486. doi: 10.1007/s00221-013-3459-7

Schmitz, L., Vesper, C., Sebanz, N., and Knoblich, G. (2018). When height carries weight: communicating hidden object properties for joint action. *Cogn. Sci.* 42, 2021–2059. doi: 10.1111/cogs.12638

Schubert, T. W., and Semin, G. R. (2009). Embodiment as a unifying perspective for psychology. *Eur. J. Soc. Psychol.* 39, 1135–1141. doi: 10.1002/ejsp.670

Sebanz, N., and Shiffrar, M. (2009). Bluffing Bodies: inferring intentions from actions. *Psychon. Bull. Rev.* 16, 170–175. doi: 10.3758/PBR.16.1.170

Sevdalis, V., and Keller, P. E. (2009). Self-recognition in the perception of actions performed in synchrony with music. *Ann. N Y Acad. Sci.* 1169, 499–502. doi: 10.1111/j.1749-6632.2009.04773.x

Sevdalis, V., and Keller, P. E. (2010). Cues for self-recognition in point-light displays of actions performed in synchrony with music. *Conscious. Cogn.* 19, 617–626. doi: 10.1016/j.concog.2010.03.017

Sevdalis, V., and Keller, P. E. (2011a). Captured by motion: dance, action understanding, and social cognition. *Brain Cogn.* 77, 231–236. doi: 10.1016/j.bandc.2011.08.005

Sevdalis, V., and Keller, P. E. (2011b). Perceiving performer identity and intended expression intensity in point-light displays of dance. *Psychol. Res.* 75, 423–434. doi: 10.1007/s00426-010-0312-5

Sevdalis, V., and Keller, P. E. (2012). Perceiving bodies in motion: expression intensity, empathy, and experience. *Exp. Brain Res.* 222, 447–453. doi: 10.1007/s00221-012-3229-y

Sevdalis, V., and Keller, P. E. (2014). Know thy sound: perceiving self and other in musical contexts. *Acta Psychol.* 152, 67–74. doi: 10.1016/j.actpsy.2014.07.002

Sevdalis, V., and Raab, M. (2014). Empathy in sports, exercise, and the performing arts. *Psychol. Sport Exerc.* 15, 173–179. doi: 10.1016/j.psychsport.2013.10.013

Sevdalis, V., and Raab, M. (2016). Individual differences in athletes' perception of expressive body movements. *Psychol. Sport Exerc.* 24, 111–117. doi: 10.1016/j.psychsport.2016.02.001

Sevdalis, V., and Wöllner, C. (2016). "Capturing motion for enhancing performance: an embodied cognition perspective on sports and the performing arts," in *Performance Psychology*, eds M. Raab, B. Lobinger, S. Hoffmann, A. Pizzera and S. Laborde (Amsterdam: Academic Press Elsevier), 223–234.

Tomasello, M. (2019). *Becoming Human: A Theory of Ontogeny.* Cambridge, MA: Harvard University Press.

Vesper, C. (2014). How to support action prediction: Evidence from human coordination tasks. *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, Edinburgh, 655–659.

Vesper, C., Abramova, E., Bütepage, J., Ciardo, F., Crossey, B., Effenberg, A., et al. (2017a). Joint action: mental representations, shared information and general mechanisms for coordinating with others. *Front. Psychol.* 7, 2039–2039. doi: 10.3389/fpsyg.2016.02039

Vesper, C., Schmitz, L., and Knoblich, G. (2017b). Modulating action duration to establish non-conventional communication. *J. Exp. Psychol. Gen.* 164, 1722–1737. doi: 10.1037/xge0000379

Vesper, C., Butterfill, S., Knoblich, G., and Sebanz, N. (2010). A minimal architecture for joint action. *Neural Netw.* 23, 998–1003. doi: 10.1016/j.neunet.2010.06.002

Vesper, C., and Richardson, M. J. (2014). Strategic communication and behavioral coupling in asymmetric joint action. *Exp. Brain Res.* 232, 2945–2956. doi: 10.1007/s00221-014-3982-1

Vesper, C., Schmitz, L., Safra, L., Sebanz, N., and Knoblich, G. (2016). The role of shared visual information for joint action coordination. *Cognition* 153, 118–123. doi: 10.1016/j.cognition.2016.05.002

Vicary, S., Sperling, M., von Zimmermann, J., Richardson, D. C., and Orgs, G. (2017). Joint action aesthetics. *PLoS One* 12:e0180101. doi: 10.1371/journal.pone.0180101

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., and Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition* 101, 80–113. doi: 10.1016/j.cognition.2005.09.003

Volman, I., Noordzij, M. L., and Toni, I. (2012). Sources of variability in human communicative skills. *Front. Hum. Neurosci.* 6:310. doi: 10.3389/fnhum.2012.00310

Williamson, R. A., and Brand, R. J. (2014). Child-directed action promotes 2-year-olds' imitation. *J. Exp. Child Psychol.* 118, 119–126. doi: 10.1016/j.jecp.2013.08.005

Williamon, A., and Davidson, J. W. (2002). Exploring co-performer communication. *Music. Sci.* 6, 53–72. doi: 10.1177/102986490200600103

Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev.* 9, 625–636. doi: 10.3758/bf03196322

Wilson, M., and Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychol. Bull.* 131, 460–473. doi: 10.1037/0033-2909.131.3.460

Wöllner, C. (2012a). Self-recognition of highly skilled actions: a study of orchestral conductors. *Conscious. Cogn.* 21, 1311–1321. doi: 10.1016/j.concog.2012.06.006

Wöllner, C. (2012b). Is empathy related to the perception of emotional expression in music? A multimodal time-series analysis. *Psychol. Aesthet. Creat. Arts* 6, 214–223. doi: 10.1037/a0027392

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Human Neuroscience

ORIGINAL RESEARCH
published: 17 July 2020
doi: 10.3389/fnhum.2020.00234

Subjective Evaluation of Performance in a Collaborative Task Is Better Predicted From Autonomic Response Than From True Achievements

Alexander Maye[1]*, *Jürgen Lorenz*[2], *Mircea Stoica*[1] *and Andreas K. Engel*[1]

[1] *Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany,*
[2] *Laboratory of Human Biology and Physiology, Faculty of Life Science, Applied Science University, Hamburg, Germany*

Edited by:
Praveen K. Pilly,
HRL Laboratories, United States

Reviewed by:
Stefano Govoni,
University of Pavia, Italy
Andrew Paul Owens,
King's College London,
United Kingdom

***Correspondence:**
Alexander Maye
a.maye@uke.de

Specialty section:
This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

Received: *18 November 2019*
Accepted: *28 May 2020*
Published: *17 July 2020*

Citation:
Maye A, Lorenz J, Stoica M and Engel AK (2020) Subjective Evaluation of Performance in a Collaborative Task Is Better Predicted From Autonomic Response Than From True Achievements. Front. Hum. Neurosci. 14:234. doi: 10.3389/fnhum.2020.00234

Whereas the fundamental role of the body in social cognition seems to be generally accepted, elucidating the bodily mechanisms associated with non-verbal communication and cooperation between two or more persons is still a challenging endeavor. In this article we propose a fresh approach for investigating the function of the autonomic nervous system that is reflected in parameters of heart rate variability, respiration, and electrodermal activity in a social setting. We analyzed autonomic parameters of dyads solving a target-tracking task together with the partner or individually. A machine classifier was trained to predict the subjects' rating of performance and collaboration either from tracking error data or from the set of autonomic parameters. When subjects collaborated, this classifier could predict the subjective performance ratings better from the autonomic response than from the objective performance of the subjects. However, when they solved the task individually, predictability from autonomic parameters dropped to the level of objective performance, indicating that subjects were more rational in rating their performance in this condition. Moreover, the model captured general knowledge about the population that allows it to predict the performance ratings of an unseen subject significantly better than chance. Our results suggest that, in particular in situations that require collaboration with others, evaluation of performance is shaped by the bodily processes that are quantified by autonomic parameters. Therefore, subjective performance assessments appear to be modulated not only by the output of a rational or discriminative system that tracks the objective performance but to a significant extent also by interoceptive processes.

Keywords: biophysical methods, self-perception, joint attention, embodied cognition, human behavior

1. INTRODUCTION

Today the idea that social cognition is not a purely mental phenomenon but also involves the body seems to be accepted by many researchers. Yet we need to better understand the bodily mechanisms during non-verbal social interaction. Relevant phenomena range from the tactile and proprioceptive perception of touches, forces, and torques that are produced by physical contact between persons either directly or mediated by objects, over observation or gestures involved in interacting across peripersonal space, to complex emotional processes that characterize shared experiences and behavior between individuals. To understand how individuals perceive other agents and control social behavior therefore requires the examination of the relationship between the autonomic nervous system and interoceptive and emotional functions.

In this respect the analysis of heart rate variability (HRV) as an autonomic indicator provides a particularly interesting approach. Normal heartbeat is automatically generated by autorhythmic cells in the sinoatrial node. The cardiac pacemaker possesses a substantial level of autonomy from the brain in that the basic activity pattern continues even when the heart is denervated. Still this activity is permanently modulated by the brainstem through the sympathetic and parasympathetic nervous system, accelerating and decelerating heart rate, respectively. We consider HRV as an index for the nervous and hormonal signals that distinctly modulate the strength of sympathetic and parasympathetic actions on the heart pacemaker, the sinoatrial node, which results in changes of inter-beat intervals (Shaffer and Ginsberg, 2017). Parasympathetic modulation of cardiac activity by the brain is mediated through the right vagus nerve; sympathetic modulation is exerted through inputs from postganglionic efferents originating in the stellate ganglion. The brain, in turn, receives afferent signals from aortic, carotide, and pulmonary baroceptors through the vagus nerve (Ellis and Thayer, 2010). Respiratory modulation of heart rate with increases during inspiration and decreases during expiration is called respiratory sinus arrhythmia (RSA, Berntson et al., 1993). Although RSA grossly reflects the rhythmical fluctuation of pulmonary vagal afferent and cardiac vagal efferent effects upon the sinoatrial node in synchrony with the breathing cycle, experimental and anatomical evidence indicate additional RSA-independent sources of cardiac vagal tone (Grossman and Taylor, 2007; Farmer et al., 2016). Together with electrodermal activity (EDA), which is typically considered as an index of sympathetic activity (Dawson et al., 2007), parameters of the respiration rhythm (RR) and HRV therefore may be considered as a complex status indicator of the autonomic nervous system.

A mutual brain-viscera interaction has been highlighted as early as in the nineteenth century (Charles et al., 1998) for the understanding of the interplay between emotion and cognition. Damasio (1999) considers visceral input as part and parcel of emotion and suggests that in particular background feelings "are a faithful index of momentary parameters of inner organism state" which have the "temporal and spatial shape of ... the striated muscle of heart and chest" as core ingredients (p. 286f.). The central autonomic network (Benarroch, 1993, 2014), linking the brainstem with forebrain structures through feed-back and feed-forward loops, is responsible for generating visceromotor, neuroendocrine, and behavioral responses that are flexibly adapted to environmental demands (Thayer and Lane, 2000). Indeed several studies have found a relation between HRV and the adaptive and functional top-down and bottom-up cognitive modulation of emotional stimuli (Park and Thayer, 2014). Since activity in anterior regions of the prefrontal cortex correlates with HRV specifically during emotionally challenging situations, individuals with high HRV may be particularly efficient in recruiting the "social cognition" network in emotional contexts (Beffara et al., 2016). Correlations of emotional state have been shown to exist with individual HRV parameters (Zhu et al., 2019), such as mean heart rate (Yoshino et al., 2011; Choi et al., 2017) or the high-frequency component of heart rate fluctuations (Lane et al., 2009), as well as with subsets of HRV parameters (Rainville et al., 2006). Moreover, higher levels of heart rate synchrony have been suggested as a marker of interpersonal trust (Mitkidis et al., 2015). Several studies revealed correlations of the emotional state with parameters of RR (Del Negro et al., 2018) and EDA (Sequeira et al., 2009). Real-time feedback about HRV coherence in pairs or groups of people is used to investigate whether learning to regulate physiological coherence helps increasing social coherence, leading to increased prosocial behaviors, improved communication, cooperation, creativity, and decision making (McCraty, 2017).

The influence of visceral information on perceptual processes and cognitive functions however is less well-explored. Recent findings show that heartbeat-evoked neural activity can modulate perceptual thresholds and shape visual conscious experience (Park et al., 2014). This led to the hypothesis that the neural representation of visceral information, projected through multiple anatomical pathways to a network of brain regions including posterior insula, ventral anterior cingulate cortex, amygdala and somatosensory cortex, constitutes an implicit frame which could explain the subjective nature of perceptual experience and link it with emotions and the notion of the self (Park and Tallon-Baudry, 2014). This hypothesis gains support from the observation that heartbeat-evoked neural responses co-vary with the self-relatedness of ongoing spontaneous thoughts (Babo-Rebelo et al., 2016). Heart rate and EDA have been found to correlate with various dimensions of the subjective experience of playing a computer game, linking quantitative parameters with the quality of user experience (Drachen et al., 2010).

The study we present in this article is geared to contribute at least two novel aspects to this interesting line of research. The first is to go beyond establishing correlative relations and explore possibilities for actually predicting the outcome of the evaluation of the subjective experience. We approach this question by training a machine classifier to predict ratings of subjective experience from autonomic parameters and analyzing the prediction performance. If the trained model performs above chance level, we conclude that the autonomic response must be informative about the result of this assessment.

The second contribution of our study to the growing knowledge about brain-visceral interaction is a fresh approach for analyzing the relation between autonomic parameters and behavioral responses. The typical approach selects a single parameter or a few and analyzes how they change between normal vs. clinical conditions (Shaffer and Ginsberg, 2017). In contrast, we here follow a strategy that is inspired by machine learning approaches. Rather than considering HRV, RR, or EDA separately and analyzing individual parameters or small subsets thereof, we conceive of autonomic parameters as a feature vector which is characteristic for the stable dynamics of the body. One advantage of this approach is that it allows us to discover patterns in the parameter set which are more complex than increases or decreases of individual parameters.

We study the interaction between autonomic state and subjective experience during a joint target tracking task. Each partner controlled one of two perpendicular axes of motion, and together they had to roll a virtual ball as close as possible to a target that moved in an oblique direction. After each trial, a subjective assessment of their own performance, the partner's performance and the collaboration was requested from both participants. We investigate a potential relation between this subjective experience of performance and parameters of HRV, RR, and EDA, and we contrast it with the actual performance measured by the tracking error. We elucidate potential differences between a collaborative condition, in which subjects jointly controlled the ball, and a condition in which they solved the task individually. We do not consider the learning process for acquiring the skill to solve task here; therefore, subjects exercised the task for several days, and we analyzed the data after performance had stabilized.

Two non-exclusive hypotheses about the basis of the processes for the subjective evaluation of performance in the task will be investigated with our approach. Hypothesis 1 (H1) entails that the assessment of performance is driven by the subject's tracking of the task performance, as indexed by an objective criterion, i.e., by evaluating the tracking error. This hypothesis follows from the assumption that the individual utilizes the recall of memorized behavioral performance parameters to retrospectively rate performance. Support for H1 would be gained from good performance of the classifier for predicting ratings from the tracking error. The main idea of hypothesis 2 (H2) is that a feeling about the own performance, rather than objective discrimination, guides the ratings in the self-assessment. We postulate that this feeling about the own performance is reflected in the autonomic response and hence consider good performance of the classifier for predicting ratings from autonomic parameters as support for H2. H2 is closely related to the idea that visceral bodily states can influence how humans perceive their own actions, whereas H1 is more compatible with the view that feedback on overt motor behavior determines this experience.

2. MATERIALS AND METHODS

2.1. Experimental Setup

The cooperative task, a dual target-tracking task, was implemented on a tablet computer (iPad2, Apple Inc.). By tilting the tablet, subjects had to move a virtual ball into the center of a moving circle. The target circle moved along a straight line at a fixed speed, but reversed its direction of movement at random intervals. The animation of the virtual ball followed the kinematic equations resulting from Newton's second law to make its behavior naturalistic.

Each player controlled only one axis of the tablet. In the collaborative condition, a single target moved on a diagonal line, and subjects cooperatively moved the ball toward the target. In the individual condition, there were two confinements along the main axes of the tablet, each containing a ball and a target, and subjects tracked the target in their respective confinement independently from the partner. **Figures 1A,B** show screen shots of the two conditions.

In order to constrain movements of the subjects, we developed a custom-made frame consisting of two armrests arranged in an L-shape, handles on each armrest, and a ball joint support for the tablet computer at the intersection of the armrests (see **Figure 1C**). Subjects grasped the handle with their right hand and extended the index finger into a thimble which was attached at each side of the tablet computer. The ball joint held the tablet's balance and allowed the subjects to tilt it along the respective main axes by lifting or lowering the index finger with minimal physical effort. As the friction of the virtual ball was low, small finger movements were sufficient to move the ball around. Subjects did not report problems with controlling the ball or fatigue.

The experimental setup and the participants were placed in an electrically and acoustically shielded chamber. Subjects were instructed to not communicate verbally or gesticulate during the experiment, i.e., during the game or when submitting ratings. Compliance with this instruction was checked by the experimenter through a camera mounted in the recording chamber. Subjects were also instructed to not rotate the tablet around the vertical axis (by moving the index finger in a plane parallel to the table), but there was no mechanism for preventing such movements. Rotations in the horizontal plane, however, could not affect the ball's movement, and we have no indication that subjects used such movements as a means of covert communication.

2.2. Subjects and Study Protocol

Twenty eight subjects participated in the study (20 females, mean age 25.18 ± 3.86 years). They were right-handed and reported to be in healthy condition. Subjects gave written informed consent before commencing the experiment. As part of the procedure for obtaining informed consent from the participants, they were instructed not to smoke, consume drugs or drink alcohol or coffee before the experiment. Apart from contraceptives, participants were free from medication. The study was approved by the ethics committee of the medical association of the city of Hamburg. The experiments were performed in accordance with the Declaration of Helsinki.

Subjects were paired in 14 dyads. In all but two dyads who were exclusively female, teams were mixed-gender on at least one of the 2 days when data were recorded. On the first day of the experiment, all participants except 4 (2 dyads) declared to never have met the respective partner before.

Each dyad exercised the task on 6 consecutive days, because we aimed at analyzing the processes when task performance was dynamically stable. On each day, they completed 7 trials in either condition (collaborative/individual). The order of the conditions was randomized. Each trial lasted for 120 s. Immediately after a trial, the experimenter requested the subjects to rate their performance by asking them:

R1 : "Please rate your own performance."

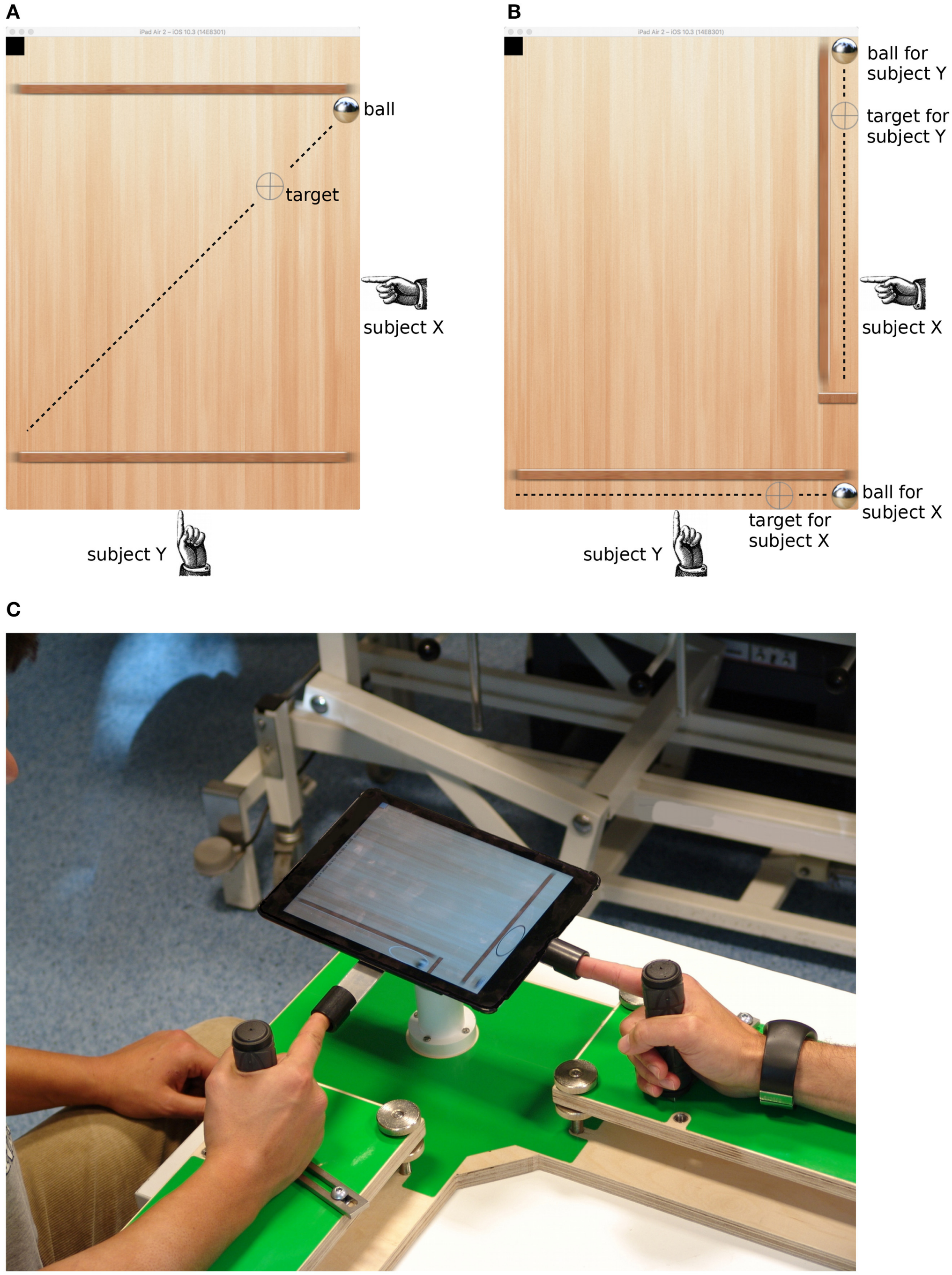


FIGURE 1 | Screenshots from the tablet computer showing the collaborative **(A)** and individual **(B)** conditions. The gray crosshair is the target that moved at constant rate along the dashed line (the line only illustrates the motion path; it was not visible to the subjects). Subjects tilted the tablet to make the ball roll to the target location. The black square in the upper left corner was used to synchronize the tablet with the amplifier and was not visible to the subjects. **(C)** Picture of the experimental setup.

R2 : "Please rate your partner's performance."
R3 : "Please rate the collaboration."

Subjects made their assessment by selecting a number between 1 and 9 (1-worst performance, 9-best) on a small remote control which they held in their left hands underneath the armrest so that the partner could not see their selection. R2 and R3 were called out only after a collaborative trial. Whereas R1 and R2 measured the subjects' impression of their performance as individuals, R3 was targeted at their performance as a team. The ratings were designed to capture different aspects in the social interaction of the subjects on the task and to facilitate correlation analyses with signatures of their body and brain activity.

On days 7 and 8, they performed the same procedure, but in addition, electrophysiological data were recorded (see below). On day 7, subjects cooperated with the same partner like the 6 days before, whereas on day 8 they were paired with a different but equally trained subject. In order to improve statistical power, but also because we here are not concerned with the differences between collaboration with a known vs. new partner, we combined the data from days 7 and 8.

2.3. Recording Physiological Signals and Calculating Autonomic Parameters

We recorded EEG, EMG, ECG, RR, and EDA simultaneously from the two subjects in each dyad using an EEG amplifier (ActiveTwo AD-box, BioSemi instrumentation) and an amplifier for physiological signals (MP35, Biopac Systems Inc.). Both amplifiers were synchronized by a common clock. Here we analyze ECG, RR, and EDA data only; results of EEG and EMG data analyses will be reported elsewhere.

The air in the recording chamber was conditioned to have a temperature of 21°C and a humidity of 40%. The interior was illuminated by 4 × 25 W LED lights on the ceiling. Except for day 8, when participants interacted with a new partner, data were recorded at the same time of the day, which differed between the teams though.

Two ECG electrodes were placed below the upper medial clavicle and on the Erb point (Eindhoven 2). ECG was sampled at 2,048 Hz. Respiration was recorded by a strain sensor on an elastic belt which subjects wore around their chest. Electrodes for recording skin conductance (EDA) were placed on the distal phalanx of the index and middle fingers on the left hand. RR and EDA were sampled at 10 Hz and a resolution of 24 bit.

ECG data analysis started by detecting R-peaks using the *qrsdetect* function in the Biosig toolbox (Vidaurre et al., 2011) in Matlab (The Mathworks). Correctness of QRS detection was checked visually for each subject. Detection of the QRS-complex resulted in so-called normal-to-normal intervals (NN) from which then HRV parameters were calculated using the *heartratevariability* function of the toolbox. Frequency-domain parameters were calculated using autoregressive modeling. Using a fast Fourier transform did not qualitatively change the results.

We used a subset of the HRV parameters that are described in (Camm et al., 1996) and, additionally, Poincaré-map parameters (SD1, SD2, r-RR; Brennan et al., 2001). **Table 1** lists all HRV parameters together with a short description of what they represent; a comprehensive explanation and their clinical relevance is given in Shaffer et al. (2014). It has to be pointed out that the HRV parameters are not independent measures of cardiac activity; rather, several of them are correlated to various degrees (Shaffer and Ginsberg, 2017).

In women, HRV parameters are known to be modulated by the phase of the menstrual cycle (Sato et al., 1995; Yildirir et al., 2001; Brar et al., 2015). There are a number of other factors, however, which also affect HRV. Age and body-mass index, for example, have been shown to exert a stronger modulation than menstrual cycle (Vallejo et al., 2005; Zhang, 2007). Likewise differences in HRV between female and male participants in our cohort had to be expected (Zhang, 2007). Since here we are not interested in the distribution of individual HRV parameters across the population, but rather in the predictive information when considered jointly, we consider menstrual cycle as one of many factors that give rise to inter-individual differences of HRV parameters and devise our methods to cope with these differences.

The respiration signal was band-pass filtered between 0.05 and 0.5 Hz, and the instantaneous breathing rate was determined from the zero-crossings of the resulting signal. Interval durations shorter than 0.5 s were considered as artifacts and removed. From the Fourier power spectrum of the filtered signal, the integrals in the frequency bands from 0.07 to 0.14 and 0.15 to 0.5 Hz, normalized by the total power, yielded spectral power features in the mid and high frequency bands, respectively (Hidalgo-Muñoz et al., 2018). RR parameters are listed in **Table 2**.

Skin conductance was decomposed into a tonic skin conductance level (SCL) and a transient skin conductance response (SCR) (Boucsein, 2010) using the continuous decomposition analysis implemented in the LEDALAB toolbox (Benedek and Kaernbach, 2010). Since SCR events may reflect stimulus-related as well as non-specific responses, and to avoid the intricacies involved in finding thresholds which define such events, we followed the approach suggested in Zhang et al. (2018)

TABLE 1 | HRV parameters.

HRV parameter	Description
ecg_mean	Mean duration of NN intervals
ecg_SDNN	Standard deviation of the NN interval
ecg_RMSSD	Square root of the mean of the squared differences between successive NNs
ecg_NN50, ecg_pNN50	Number of pairs of successive NNs that differ by more than 50 ms (NN50count) and its ratio to the total number of intervals (pNN50)
ecg_SD1, ecg_SD2	Width and length of the Poincaré plot
ecg_r_RR	Correlation coefficient in the Poincaré plot
ecg_VLF, ecg_LF, ecg_HF, ecg_tot_pwr	Power in three frequency bands (0.009–0.04, 0.04–0.15, 0.15–0.4 Hz) and total power
ecg_LF/HF	ratio of LF to HF power
ecg_LFnu, ecg_HFnu	LF and HF in normalized units, i.e., the relative value in proportion to the TotalPower minus VLF

TABLE 2 | RR parameters.

RR parameter	Description
resp_mean	Mean breathing rate
resp_std	Standard deviation of the breathing rate
resp_MF	Spectral power in the middle frequency band (0.07–0.14 Hz)
resp_HF	Spectral power in the high frequency band (0.15–0.5 Hz)

and considered the integrated SCR (iSCR) that was calculated by integrating the SCR time courses across 10 s non-overlapping time windows. SCL was treated in the same way. The means and standard deviations across the trial yielded the EDA parameters listed in **Table 3**.

For the statistical analysis of correlations between ratings and autonomic parameters, we assumed a significance threshold of 0.05 and used the false discovery rate (FDR, Benjamini and Hochberg, 1995) to correct for multiple comparisons.

2.4. Classification

We employed quadratic discriminant analysis (QDA) as a model for the relation between autonomic parameters or objective performance and ratings. QDA is a variant of linear discriminant analysis (LDA Rao, 1948; Hastie et al., 2001) which allows for different covariance matrices for each class and hence more complex decision boundaries. Accuracies from a linear classifier were always lower, which relates to the finding that the inclusion of quadratic terms seemed to improve model accuracy (Beffara et al., 2016). Although the problem naturally is one of ordinal classification or regression, none of the corresponding methods that we tested (LASSO, random forests, support vector machines) outperformed QDA. We used the implementations of QDA and LDA that are provided in Matlab (*classify*) and custom scripts.

Models were trained on pooled data from all subjects ($N = 28$ subjects $\times$ 14 trials $\times$ 2 days $=$ 784 samples). In order to combine autonomic data from all subjects, we normalized data by linearly mapping them to the interval $[0, 1]$. We also tried z-scores for normalizing the distribution of each parameter per subject to have zero mean and unit variance, but this did not qualitatively affect the results.

Ratings were not normalized, but we checked for outliers of the average ratings per subject with respect to the whole population. If the median of the ratings from a subject was more than 1.5 interquartile ranges above the upper quartile or below the lower quartile, we corrected ratings from this subject by subtracting the difference between the subject's median rating and the median of all other subjects' median ratings. Ratings of two subjects were corrected in this way. In order to equalize the number rating levels across the three ratings R1-3 and two conditions (collaborative/individual), and to eliminate rating levels with an insufficient number of samples for classification, we used only those trials where the rating was among the six most frequent levels. This resulted in discarding 2/6/5 trials from R1/2/3 in the collaborative condition and 3 trials from R1 in the individual condition. The resulting frequencies of the $L = 6$ rating levels are shown in **Figure 2B**.

Objective task performance was measured by calculating the tracking error. To evaluate the performance as a team, the cumulative Euclidean distance (in pixels) between the target and the ball across the duration of the trial was calculated. The individual objective performances were given by the cumulative distance (in pixels) in x-/y-direction between the target and the ball across the trial duration.

Two sets of classifiers were trained to predict ratings. One set was trained on the tracking errors. For each rating category R1–3, the corresponding tracking error was used as a measure of the subject's or the dyad's objective performance: Ratings of own performance (R1) were predicted from the tracking error along the axis that the subject controlled; ratings of the partner's performance (R2) were predicted from the tracking error along the axis that the partner controlled; and ratings of collaboration

TABLE 3 | EDA parameters.

EDA parameter	Description
iscl_mean	Mean of the integrated SCL (integration across 10 s non-overlapping time windows)
iscl_std	Standard deviation of the integrated SCL
iscr_mean	Mean of the integrated SCR (integration across 10 s non-overlapping time windows)
iscr_std	Standard deviation of the integrated SCR

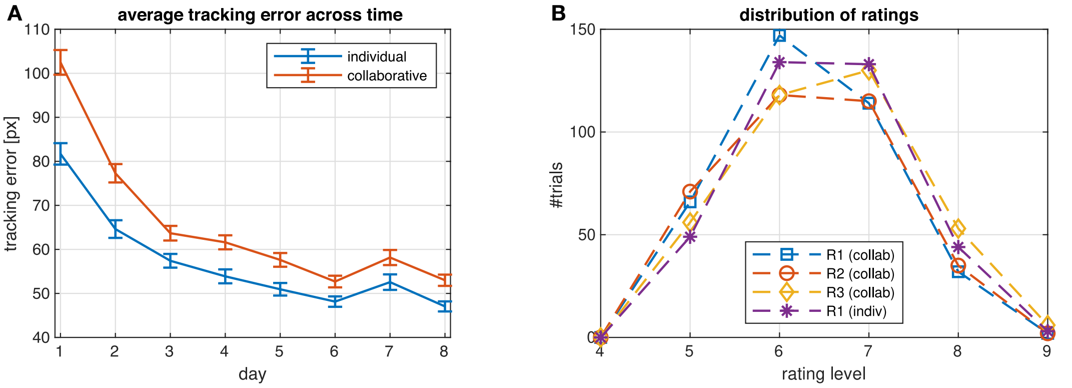


FIGURE 2 | (A) Average tracking error across all subjects for each day of participating in the study. ECG was recorded on days 7 and 8, RR and EDA on every day. **(B)** Distribution of ratings R1-3 from all subjects in the two conditions (collaborative/individual) on days 7 and 8 after correcting outliers and eliminating rating levels with an insufficient number of samples.

(R3) were predicted from the Euclidean distance between the ball and the target. The other set of classifiers was trained on the autonomic parameters.

Model performance was calculated by leave-one-sample-out cross-validation: One sample was selected from the data set, and the classifier was trained on the remaining samples. The trained classifier was then used to predict the rating of the selected sample, and the output was compared with the observed ratings. Repetition of this procedure for each sample in the data set yielded an estimate of the model's performance. To corroborate the results, and to investigate how the model responds to data from an unseen subject, we also employed leave-one-subject-out cross-validation: Here the classifier was trained on data from all but one subjects and tested on data from this subject. As this classifier captures common properties across the cohort of participants, we call this classifier the population model.

In order to reduce the redundancy between the numerous autonomic parameters and thereby improve classification performance, we employed backward feature selection for each of the three ratings and two conditions. Successively, each parameter was temporarily omitted from the data set and the resulting classification performance was determined using leave-one-sample-out cross-validation. The parameter that, when omitted from the data set, yielded the strongest performance increase was then permanently removed, and the procedure was repeated until no further performance improvements were achieved. The reduced set of parameters was then used to determine the final model performance.

Importance of the remaining parameters was estimated by permuting them individually and gauging the decrease in model performance. Parameters that cause large decreases under permutation can be considered more important than those with smaller decreases (Breiman, 2001). A data set with a permuted parameter was generated by replacing in each of the N samples of the original data set the value of the respective parameter with the values in all other samples, yielding a new data set of size $N(N-1)$. Model performance on this data set was then evaluated using leave-one-sample-out cross-validation, and the difference to the model performance on the original data set was taken as a measure of parameter importance.

Since the number of samples for each rating level was far from equal (see **Figure 2B**), we report model performance in terms of F1-scores rather than prediction accuracies. The F1-score is the harmonic mean of *recall* and *precision* of a classifier,

$$F_1 = \frac{2}{\frac{1}{recall} + \frac{1}{precision}},$$

whereby

$$recall = \frac{1}{L}\sum_{l=1}^{L}\frac{TP_l}{P_l} \qquad \text{and} \qquad precision = \frac{1}{L}\sum_{l=1}^{L}\frac{TP_l}{TP_l + FP_l}$$

are the averages of the class-specific recall and precision which are calculated from the number of true positive (TP), total positive (P), and false positive (FP) classifications.

The cross-validation methods yielded a single F1-score per condition and rating. We assessed the likelihood of obtaining the reported F1-scores by chance by running permutation tests on all classifiers (Good, 2013). In each of the 1,000 repetitions, we trained and tested the classifier on a data set in which the structure had been destroyed by randomly permuting the class labels.

3. RESULTS

3.1. Correlation Analysis of Behavioral Data

Exercising the task every day, subjects continuously improved their performance, reflected in a monotonic decrease of the average tracking error shown in **Figure 2A**. The only exception was on day 7, when the introduction of ECG and EEG recording likely affected the experimental routine acquired during the previous days, leading to a transient decrease of task performance. Average performance was higher when subjects tracked individual targets compared to when they collaborated to track the target jointly. On days 7 and 8 the average tracking error was 49.8/55.6 pixels in the individual/collaborative condition, respectively ($p = 2.4466e^{-4}$, paired two-sided t-test). The subjects rated their own performance slightly higher in the individual than in the collaborative condition (6.3269 vs. 6.1319, $p = 0.0429$).

Participants mainly used the upper half of the 9-point scale (values 5–9) for assessing task performances, with 6 and 7 being the most frequent responses (**Figure 2B**). The distributions of the responses was nearly normal (Lilliefors test, p-values for the four ratings between 0.01 and 0.021).

We analyzed the relation between the performance assessments of the two partners in the collaborative condition by calculating Pearson correlation coefficients between all possible combinations of the ratings R1–3. For most rating combinations, correlations were more or less evenly distributed across the negative and positive ranges (**Figure 3**). Only correlations between mutual ratings of the partner's performance (R2 subject x vs. R2 subject y, center panel) and between ratings of the partner's performance and collaboration (R2 subject x vs. R3 subject y, middle panel in the bottom line) appeared to be significant (medians different from zero: $p = 0.013$ and $p = 0.015$, respectively, Wilcoxon signed rank test). Despite this prevalence of positive correlations in the mutual assessment of the partner's performance, the rather flat distribution of correlations between ratings of the collaboration of the partners (R3 subject × vs. R3 subject y, lower right panel) indicates that in most dyads, a feeling of 'good collaboration' was rarely reciprocated by the partner, and in some dyads this feeling even was inverse.

Correlations were weak in general; only in a few dyads they reached a threshold of $p < 0.05$ (FDR-corrected) which is marked by the dashed lines in **Figure 3**. This suggests that subjects in a dyad rated the performance independently of the partner. We therefore investigated whether there was a systematic relation between the ratings within each subject instead. In contrast to the relation between the ratings in the dyad, most ratings from an individual subject were positively correlated (median different

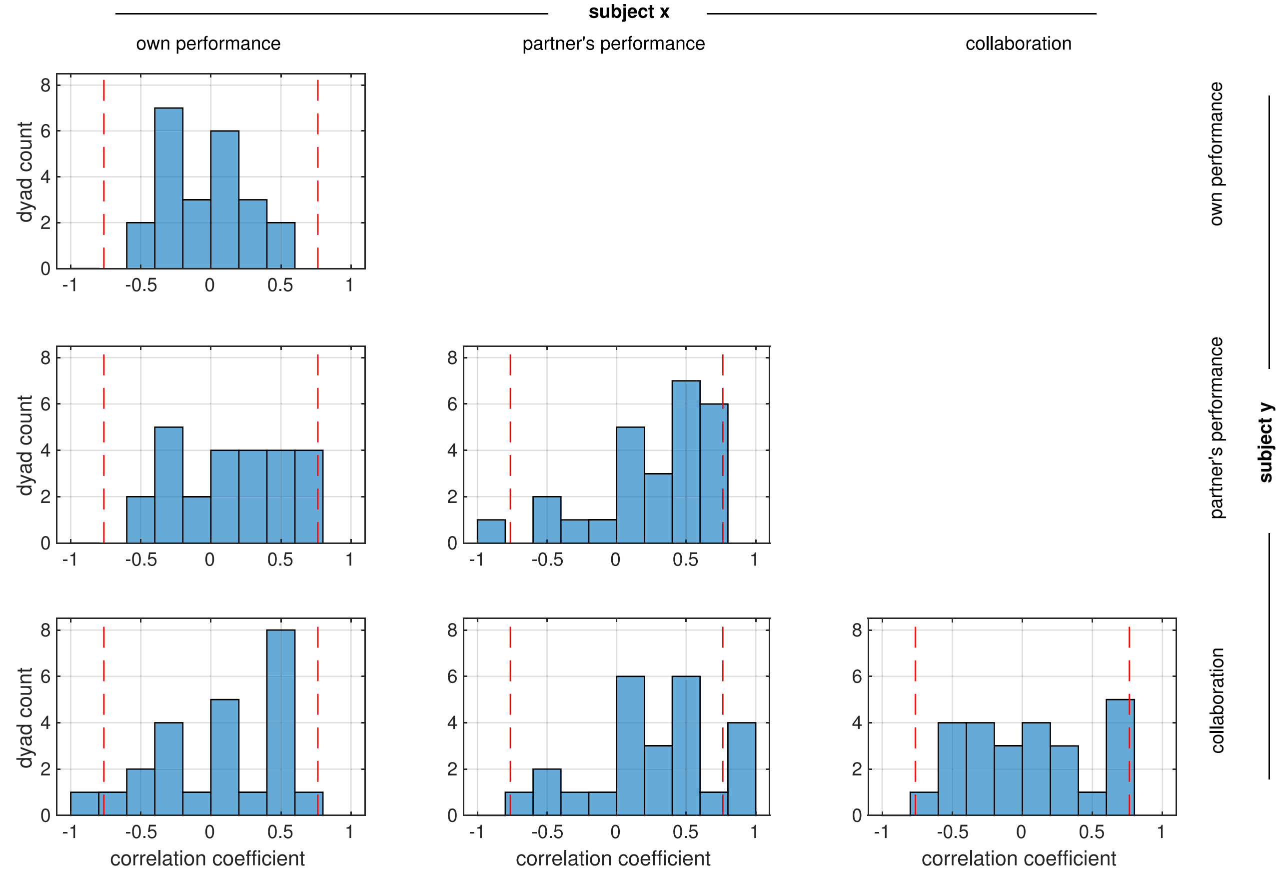


FIGURE 3 | Distribution of correlation coefficients between the ratings from the two partners in each dyad. All possible correlations were calculated, e.g., the top panel shows the correlation coefficients between the ratings of own performance (R1) from the two partners. Red dashed lines mark the strength beyond which correlations are significant ($p < 0.05$, FDR-corrected).

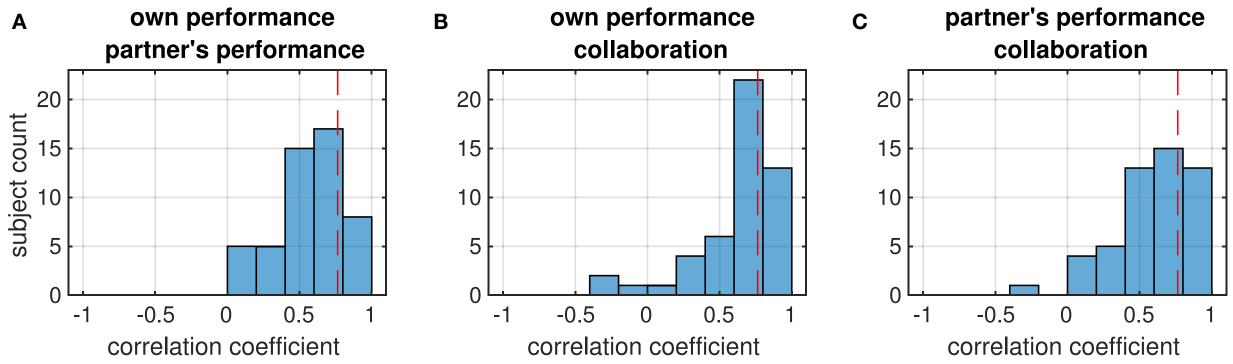


FIGURE 4 | Distribution of correlation coefficients between the ratings (printed above each histogram) from each subject. Red dashed lines mark the strength beyond which correlations are significant ($p < 0.05$, FDR-corrected). The four panels show the distribution of correlations between **(A)** ratings of the own and the partner's performance on the own axis (collaborative condition), **(B)** the partner's performance rating and the tracking error on the partner's axis, **(C)** ratings of collaboration and the Euclidean tracking error, **(D)** own performance rating and tracking error on the own axis (individual condition).

from zero: all $p < 1e^{-8}$, Wilcoxon signed rank test), reaching a significance threshold of $p < 0.05$ (FDR-corrected) in several subjects (**Figure 4**).

A third set of correlation analyses was run to figure out in how far the subjects' ratings reflected actual task performance. From the trials in the collaborative condition, we calculated the correlation of the own performance ratings with the tracking error along the subject's axis, the partner's performance rating with the tracking error along the partner's axis, and the collaboration rating with the absolute tracking error. Whereas most subjects showed a negative correlation between the tracking error and their ratings (median different from zero: all $p < 1e^{-5}$, Wilcoxon signed rank test), only in a few of them this correlation was significant, suggesting that most subjects were hardly objective in the assessments of their own performance, their partner's performance or the success of their collaboration (**Figure 5**). In addition, we could not observe any significant difference in the distribution of correlation coefficients between the four ratings (all $p > 0.82$, Wilcoxon signed rank test).

3.2. Analyzing the Relation Between Ratings and Individual Autonomic Parameters

Following a conventional approach for studying the properties of autonomic parameters in cognitive tasks, we analyzed correlations between each of the 23 parameters that were used in this study (see **Tables 1–3**) and the ratings. From the set of HRV parameters, one ore more were significantly correlated with each of the four different ratings (see **Figure 6**). In particular the r_RR parameter (correlation coefficient in the Poincaré plot) correlated with ratings of own performance in the collaborative and individual conditions as well as with ratings of collaboration. In the individual condition, 8 of the 15 HRV parameters were correlated with ratings of own performance, whereas for the collaborative condition, only 3 or less parameters correlated with the ratings. Collaboration was the only rating that correlated with one of the EDA parameters, the standard deviation of the skin conductance level (iscl_std). Own performance in the collaborative condition was the only rating that correlated with one of the RR parameters, the mean respiration rate (resp_mean). About half of the parameters however did not show significant correlations with at least one of the four ratings.

3.3. Predicting Ratings From Autonomic Parameters and Objective Performance

The correlation analysis in the previous section showed that some of the autonomic parameters bore a linear relation to at least one of the ratings. The correlations were calculated across all participants and trials. In the next step we explored to what extent this relation would enable a machine classifier to make predictions about single trials. We successively trained a classifier to predict ratings from each autonomic parameter individually and evaluated the performance by leave-one-sample-out cross-validation. Since samples in the data set were unevenly distributed across rating levels 5–9 (see **Figure 2B**), we could not employ accuracy for quantifying classification performance. For non-equally distributed target classes, classification accuracy may be a misleading quality measure, because a classifier could achieve high accuracy values by simply deciding for the most frequent class. Instead, we used F1-scores to compare the classification performances on different autonomic parameters. We determined the likelihood of observing these performance values when in fact there is no structure in the data by comparing them against the distribution of F1-scores on surrogate data. The result is shown in **Figure 7**. None of classification performances exceeded the chance level (all $p > 0.05$, FDR-corrected), that is, none of the ratings could be predicted from any of the autonomic parameters.

To investigate whether this situation could be changed when autonomic parameters are considered jointly rather than individually, we trained another set of models on feature vectors which were composed from subsets of the autonomic parameters. Starting from a feature set with all parameters, we eliminated one by one until classification performance did not improve further (backward feature selection). The classification results on the optimized parameter set are summarized in **Figure 8**. Prediction performance from the aggregated parameters was well above chance level for all ratings. Prediction of ratings was generally better in the collaborative than in the individual condition. To explore the stability of these findings, we also tested whether the models captured some general properties of the relation between ratings and the autonomic response across

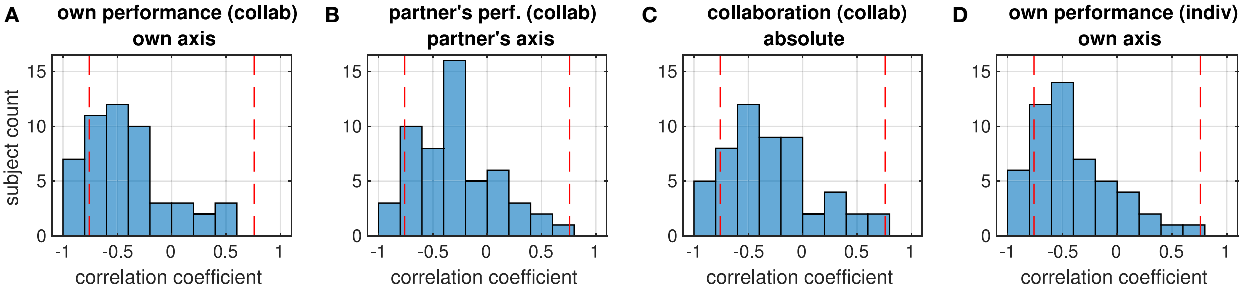


FIGURE 5 | Distribution of correlation coefficients between ratings (given in the first line of the title) and objective performance (tracking error, second title line) of all subjects. Red dashed lines mark the strength beyond which correlations are significant ($p < 0.05$, FDR-corrected). Correlations are mostly negative because subjects were asked to rate higher performance, which corresponds to smaller tracking errors, by larger values. The four panels show the distribution of correlations between **(A)** own performance rating and tracking error on the own axis (collaborative condition), **(B)** the partner's performance rating and the tracking error on the partner's axis, **(C)** ratings of collaboration and the Euclidean tracking error, **(D)** own performance rating and tracking error on the own axis (individual condition).

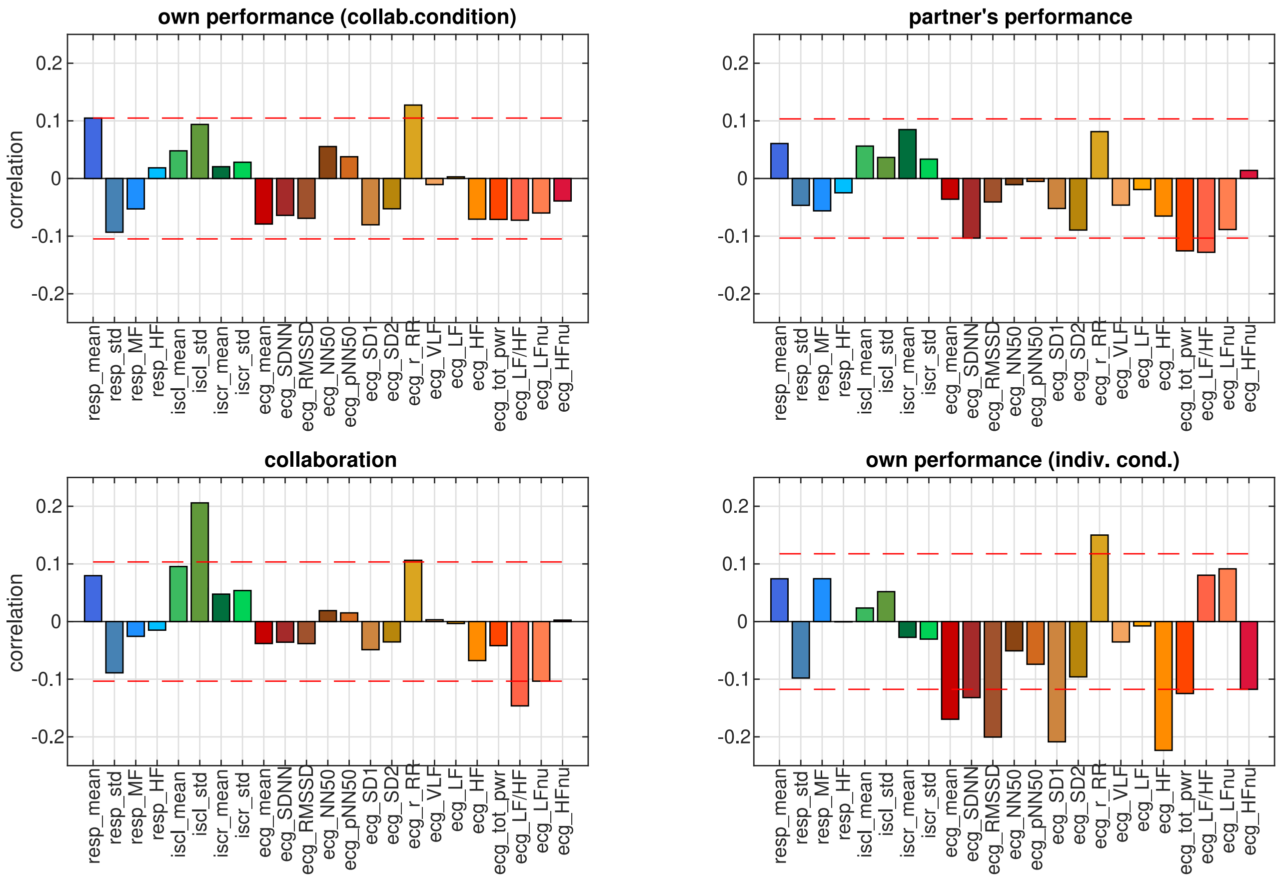


FIGURE 6 | Correlation coefficients between individual autonomic parameters and ratings. RR parameters have blue hues, EDA parameters green colors, and HRV parameters shades of red. Dashed red lines visualize the strength beyond which correlations are significant ($p < 0.05$, FDR-corrected).

the population, which would allow it to make predictions about unseen subjects. We therefore ran a leave-one-*subject*-out cross-validation on the same parameter sets. Prediction performances were generally lower under this cross-validation method, which had to be expected since generalizing to a unseen subject is harder than generalizing to new samples when data from all subjects were already seen in the training. Nevertheless, this analysis revealed that the subjective performance evaluation of an unseen participant in the collaborative condition could still be predicted from the autonomic response of this participant above chance level, whereas this was not possible when the participants solved the task individually.

For the relation of the objective task performance measured by the tracking error to the subjective experience, the analysis in section 3.1 showed that correlations were significant only in a few participants (c.f. **Figure 5**), but that for most of them, the correlations were stronger than those for individual autonomic parameters (shown in **Figure 6**). This raised the question whether these stronger correlations could result in a better predictability of the subjective performance evaluation from the objective task performance than from individual autonomic parameters. The results of predicting ratings from the objective performance using both cross-validation methods are shown in **Figure 8**. Indeed, prediction performance of own performance ratings from the tracking error reached the chance level, which is a clear improvement compared to the prediction from individual autonomic parameters (c.f. **Figure 7**). However, when subjects collaborated, predicting ratings from objective performance was always inferior to the prediction from aggregated autonomic parameters. In contrast, the generalization capabilities of the model across subjects seemed to be better for the tracking error than for the autonomic response when subjects tracked targets individually. Numerical values of F1 scores and *p*-values are listed in **Tables 4**, **5**.

3.4. Analyzing the Relevance of Individual Parameters

Finally, we were interested in the importance of individual autonomic parameters in the optimized feature set for the prediction performance. We therefore permuted each parameter and ordered them according to the incurred decrease in the model's prediction performance (**Figure 9**). Comparing the three groups of autonomic parameters, we observed that all RR parameters are among the seven most important parameters for the prediction of own performance in the collaborative condition, whereas they rank lower in predicting the remaining

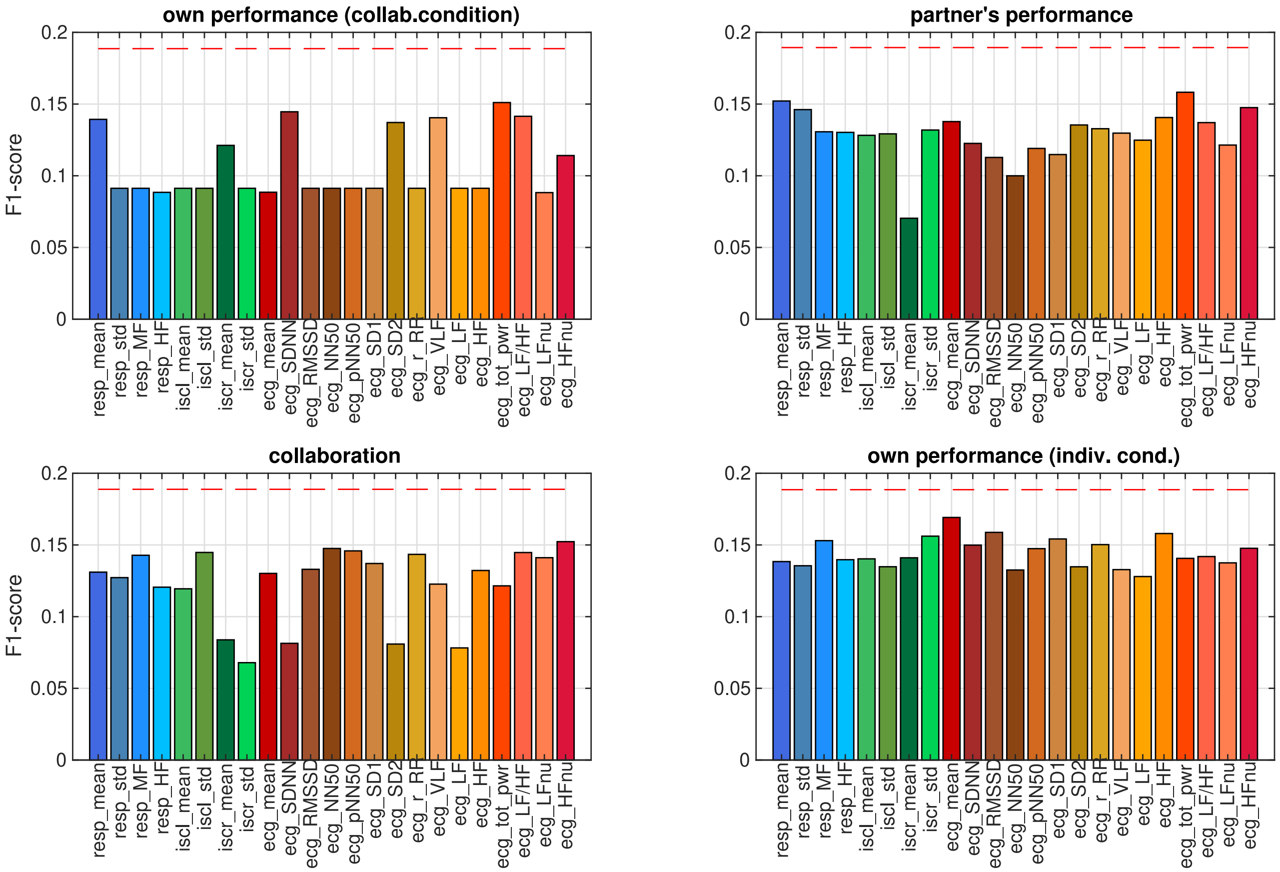


FIGURE 7 | F1-scores of predicting ratings from individual autonomic parameters, estimated using leave-one-sample-out cross-validation. Dashed red lines mark the 0.05 quantile of obtaining the corresponding F1-score or higher by chance (FDR-corrected).

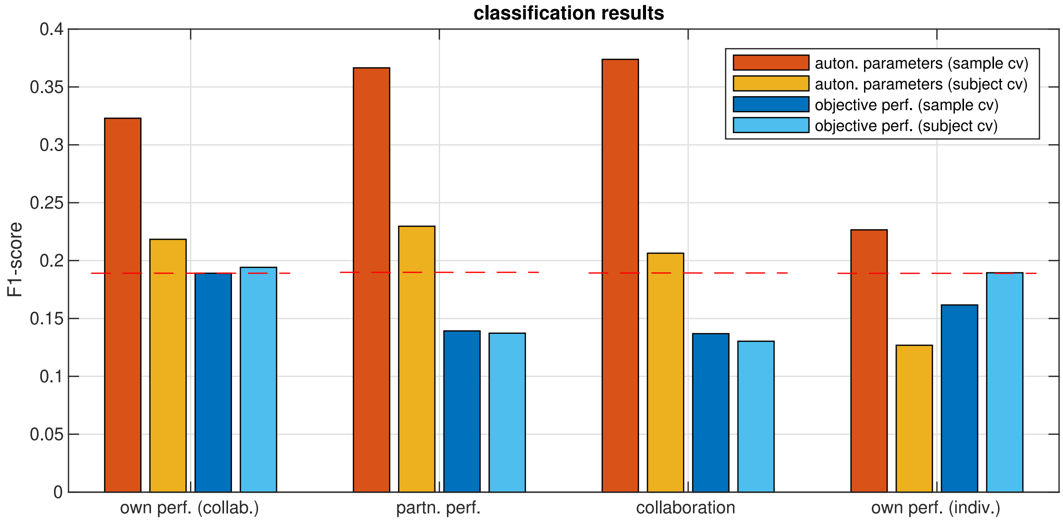


FIGURE 8 | F1-scores of predicting ratings from aggregated autonomic parameters and from objective performance, estimated using cross-validation on a single sample (sample cv) and on all samples from one subject (subject cv). Dashed red lines mark the 0.05 quantile of obtaining the corresponding F1-score or higher by chance (FDR-corrected).

TABLE 4 | F1-scores, precision, and recall from leave-one-sample-out cross-validation and *p*-values from a randomization test.

Rating	Classification from auton. param.				Classification from obj. perform.			
	F1-score	*p*	prec.	recall	F1-score	*p*	prec.	recall
Own performance (R1, collab.)	0.32	0	0.43	0.26	0.19	0.048	0.18	0.19
Partner performance (R2)	0.37	0	0.47	0.3	0.14	0.866	0.11	0.18
Collaboration (R3)	0.37	0	0.46	0.31	0.14	0.905	0.11	0.18
Own performance (R1, indiv.)	0.23	0	0.23	0.23	0.16	0.461	0.14	0.2

TABLE 5 | F1-scores, precision, and recall from leave-one-subject-out cross-validation and *p*-values from a randomization test.

Rating	Classification from auton. param.				Classification from obj. perform.			
	F1-score	*p*	prec.	recall	F1-score	*p*	prec.	recall
Own performance (R1, collab.)	0.22	0.001	0.22	0.22	0.19	0.027	0.19	0.2
Partner performance (R2)	0.23	0	0.25	0.22	0.14	0.889	0.11	0.18
Collaboration (R3)	0.21	0.008	0.18	0.24	0.13	0.96	0.11	0.17
Own performance (R1, indiv.)	0.13	0.984	0.14	0.11	0.19	0.045	0.18	0.2

ratings (with the exception of the mean breathing rate for predicting the partner's performance). EDA parameters play a role in predicting the partner's performance and collaboration but are less important in predicting own performance. For all ratings, different combinations of HRV parameters have the strongest influence on the prediction performance. From the set of HRV parameters, frequency-related parameters (HF, HFnu, LF, LF/HF, VLF) seem to be critically involved in the prediction of own and partner performance in the collaborative condition, whereas prediction of collaboration and own performance in the individual condition relies more on parameters which capture the regularity of the heart beat intervals in the time domain (NN50, pNN50, SD1, SD2).

4. DISCUSSION

This study investigated the relation between the subjective assessment of performance and activity of the autonomic nervous system indexed by HRV, RR, and EDA in a joint target-tracking task. Ratings of collaboration, partner and own performance, which were reported by the subjects after each trial, were highly correlated within each subject. This indicates that subjects did not differentiate much between the individual contributions inquired by the different questions and possibly rated them on the basis of a general impression of the success in tracking the target instead. Partners within a dyad however rarely agreed upon the success of their collaboration or the performance of their partner. Hence, it seems that the experience of joint task performance was not shared among the partners. Whereas the joint target-tracking task and the instructions engaged many of the coordinating mechanisms that constitute joint action (Vesper et al., 2017) (e.g., monitoring, joint attention and shared gaze, haptic coupling, emotion understanding, and expression), critical components for making the task completion a joint experience may have been missing. The independent control of orthogonal axes was likely to impact the joint action goal as well as the task co-representation, leading to a collaboration experience that was not systematically reciprocated. Yet subjects reacted differently to the collaborative and to the individual condition, which became evident in the lower tracking error and higher ratings of their own performance in the individual condition. The current results should therefore be interpreted on the background of an experimental manipulation of joint attention (Maye et al., 2017) rather than joint action.

We compared the capability of predicting ratings from aggregated autonomic parameters and from the tracking error as an objective measure of task performance. We found that predictions from autonomic parameters generally were more reliable, in particular for ratings of the partner's performance and the success of the collaboration. Autonomic parameters were also more effective in predicting ratings of own performance, but the prediction performance was lower for trials in which the target was tracked individually than for joint tracking. The prediction performance for the own performance from the tracking error seemed to be less affected by the condition though. Taken together, our results suggests that the autonomic response is more informative for inferring the subjective experience in the collaborative condition and less so when interaction with the partner is not required. For the efficiency of the objective task performance, however, the difference is between assessments of own performance, no matter whether in a collaborative or individual context, and the evaluation of the contribution of the partner to achieving a common goal. These findings may provide support for efforts to increase social coherence by using realtime-feedback for enabling group members to co-regulate HRV coherency (McCraty, 2017).

With respect to the two hypotheses about the origin of the subjective performance assessment, our results suggest that H1 can not sufficiently well explain how subjects arrived at their ratings. H2 entails that the subjects' ratings were guided by a

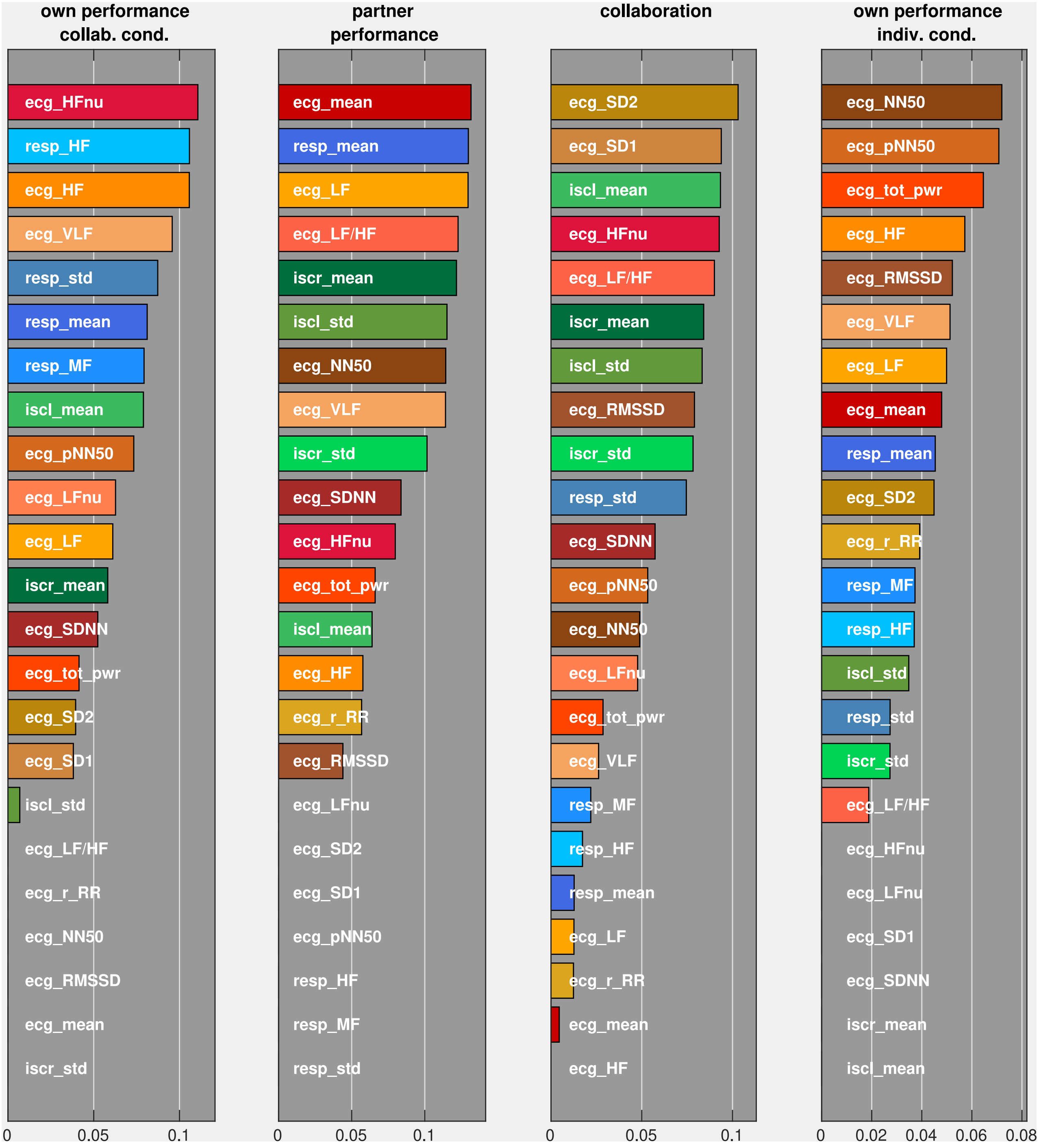


FIGURE 9 | Average decrease of F1-scores (on the abscissa) under permutations of individual predictors (on the ordinate). Predictors were ordered according to their associated F1-score change.

feeling about the performance, and that this feeling is modulated by visceral information. In particular in a collaborative setting, the autonomic response may be a good indicator for the outcome of this subjective performance assessment. The stronger coupling between ratings and these parameters in the collaborative condition suggests that subjects rely more on the internal state of their body when assessing the outcome in a collaborative task.

A crucial point of this study is that the relation between the autonomic response and the outcome of a subjective assessment could only be observed when the parameters were considered in an integrated manner. Traditional studies investigate the relation between a single (or a few) parameter(s) and the experimental paradigm. Quite often, these studies discriminate between fluctuations on a short time scale which are reflected in short-term components of the HRV (e.g., RMSSD, NN50count, SD1, or HF) and variability on a longer time scale as indicated by long-term components (e.g., SDNN, SD2, or LF), and interpret effects as shifts of the sympatho-vagal balance in the modulation of cardiac activity. This approach is followed on the background that long-term components, such as LF, and short-term components, such as HF, reflect primarily sympathetic and vagal modulations, respectively. Here, this approach did not reveal such an obvious systematic relation between any of the typical HRV parameters (Camm et al., 1996; Brennan et al., 2001) and ratings. However, considering the same parameters as elements of a feature vector and training a simple classifier allowed us to obtain a population model which predicted the result of the subjective assessment better than that of a model based on the objective performance. This indicates that there is a systematic relation between autonomic parameters and subjective performance evaluation that is shared across the population. Since we observed this relation only for aggregated parameters and not for individual ones, we cannot interpret the result with respect to only one of the mental-cognitive phenomena that are typically considered in the literature in relation to autonomic parameters. Yet, one should consider that sympathetic and parasympathetic activity are not always entirely antagonistic, rendering a simple concept of sympathetic-parasympathetic balance inadequate in complex cognitive, emotional, and behavioral situations.

Whereas, the results of our study show some potential for a better understanding of the embodied nature of subjective experience, the ramifications of the approach have to be elucidated in future investigations. An important issue in this respect is the demography of the cohort and the composition of the dyads. Participants in our study were young students, and a clear majority of them were females. As gender is known to affect the dynamics of autonomic parameters, it would certainly be interesting to find out whether and how this might impact the prediction capabilities of the proposed method. Other factors in this context which require systematic investigation are whether the partners in a team are from the same or different sex as well as their relationship. Another avenue for future research is probably the question in how far the findings in our study can be generalized across different tasks. Methodological difficulties notwithstanding, we think that the evidence for the importance of bodily signals in the emergence of subjective experience that we found in our study suggests that tasks which require more physical play than just flicking a finger may result in better prediction capabilities.

5. CONCLUSION

Existing studies have suggested a variety of interactions between cognitive and perceptual processes and individual autonomic parameters, but almost all of them concluded that the true relations are likely more complex. The machine-learning-inspired approach we suggest here may pave the way to understand such complex relationships. Our study suggests that the physiological activity indexed by autonomic parameters bears a relation to the subjective performance evaluation that can be stronger than that of the actual performance. This underlines the importance of considering bodily processes for understanding the mechanisms of social cognition.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ethics committee of the medical association of the city of Hamburg. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

The experiments were conceived and designed by AE, MS, and AM. The experiments were performed by MS and AM. The data were analyzed by AM and JL. The paper was written by AM, JL, and AE. All authors contributed to the article and approved the submitted version.

FUNDING

This work described in this paper was supported by the EU through project socSMCs (H2020–641321) and by the DFG through project TRR 169/B1.

ACKNOWLEDGMENTS

We would like to thank Till Schneider and Michael Hauck for insightful discussions about statistical methods and HRV parameters.

REFERENCES

Babo-Rebelo, M., Richter, C. G., and Tallon-Baudry, C. (2016). Neural responses to heartbeats in the default network encode the self in spontaneous thoughts. *J. Neurosci*. 36, 7829–7840. doi: 10.1523/JNEUROSCI.0262-16.2016

Beffara, B., Bret, A. G., Vermeulen, N., and Mermillod, M. (2016). Resting high frequency heart rate variability selectively predicts cooperative behavior. *Physiol. Behav*. 164, 417–428. doi: 10.1016/j.physbeh.2016.06.011

Benarroch, E. E. (1993). The central autonomic network: Functional organization, dysfunction, and perspective. *Mayo Clin. Proc*. 68, 988–1001. doi: 10.1016/S0025-6196(12)62272-1

Benarroch, E. E. (2014). *Central Autonomic Network*. Oxford: Oxford University Press. doi: 10.1093/med/9780199920198.003.0001

Benedek, M., and Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *J. Neurosci. Methods* 190, 80–91. doi: 10.1016/j.jneumeth.2010.04.028

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Berntson, G. G., Cacioppo, J. T., and Quigley, K. S. (1993). Respiratory sinus arrhythmia: autonomic origins, physiological mechanisms, and psychophysiological implications. *Psychophysiology* 30, 183–196. doi: 10.1111/j.1469-8986.1993.tb01731.x

Boucsein, W. (2010). *Electrodermal Activity*. New York, NY; Dordrecht; Heidelberg; London: Springer Science & Business Media. doi: 10.1007/978-1-4614-1126-0

Brar, T. K., Singh, K., and Kumar, A. (2015). Effect of different phases of menstrual cycle on heart rate variability (HRV). *J. Clin. Diagn Res*. 9:CC01. doi: 10.7860/JCDR/2015/13795.6592

Breiman, L. (2001). Random forests. *Mach. Learn*. 45, 5–32. doi: 10.1023/A:1010933404324

Brennan, M., Palaniswami, M., and Kamen, P. (2001). Do existing measures of poincare plot geometry reflect nonlinear features of heart rate variability? *IEEE Trans. Biomed. Eng*. 48, 1342–1347. doi: 10.1109/10.959330

Camm, A. J., Malik, M., Bigger, J., Breithardt, G., Cerutti, S., Cohen, R., et al. (1996). Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of pacing and electrophysiology. *Circulation* 93, 1043–1065. doi: 10.1161/01.CIR.93.5.1043

Charles, D., Paul, E., and Phillip, P. (1998) *The Expression of the Emotions in Man and Animals, 3rd Edn*. London: Harper Collins.

Choi, K.-H., Kim, J., Kwon, O. S., Kim, M. J., Ryu, Y. H., and Park, J.-E. (2017). Is heart rate variability (HRV) an adequate tool for evaluating human emotions? A focus on the use of the International Affective Picture System (IAPS). *Psychiatry Res*. 251, 192–196. doi: 10.1016/j.psychres.2017.02.025

Damasio, A. R. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Boston: Mariner Books.

Dawson, M. E., Schell, A. M., and Filion, D. L. (2007). "The electrodermal system," in *Handbook of Psychophysiology*, eds J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson (Cambridge: Cambridge University Press), 159–181. doi: 10.1017/CBO9780511546396.007

Del Negro, C. A., Funk, G. D., and Feldman, J. L. (2018). Breathing matters. *Nat. Rev. Neurosci*. 19, 351–367. doi: 10.1038/s41583-018-0003-6

Drachen, A., Nacke, L. E., Yannakakis, G., and Pedersen, A. L. (2010). "Correlation between heart rate, electrodermal activity and player experience in first-person shooter games," in *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games, Sandbox '10* (New York, NY: ACM), 49–54. doi: 10.1145/1836135.1836143

Ellis, R. J., and Thayer, J. F. (2010). Music and autonomic nervous system (dys) function. *Mus. Percept*. 27, 317–326. doi: 10.1525/mp.2010.27.4.317

Farmer, D. G., Dutschmann, M., Paton, J. F., Pickering, A. E., and McAllen, R. M. (2016). Brainstem sources of cardiac vagal tone and respiratory sinus arrhythmia. *J. Physiol*. 594, 7249–7265. doi: 10.1113/JP273164

Good, P. (2013). *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*. New York, NY: Springer Science & Business Media.

Grossman, P., and Taylor, E. W. (2007). Toward understanding respiratory sinus arrhythmia: relations to cardiac vagal tone, evolution and biobehavioral functions. *Biol. Psychol*. 74, 263–285. doi: 10.1016/j.biopsycho.2005.11.014

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning*. Springer Series in Statistics. New York, NY: Springer New York Inc. doi: 10.1007/978-0-387-21606-5

Hidalgo-Mu noz, A. R., Béquet, A. J., Astier-Juvenon, M., Pépin, G., Fort, A., Jallais, C., et al. (2018). Respiration and heart rate modulation due to competing cognitive tasks while driving. *Front. Hum. Neurosci*. 12:525. doi: 10.3389/fnhum.2018.00525

Lane, R. D., McRae, K., Reiman, E. M., Chen, K., Ahern, G. L., and Thayer, J. F. (2009). Neural correlates of heart rate variability during emotion. *NeuroImage* 44, 213–222. doi: 10.1016/j.neuroimage.2008.07.056

Maye, A., Isern-Mas, C., Barone, P., and Michael, J. A. (2017). Sensorimotor accounts of joint attention. *Scholarpedia* 12:42361. doi: 10.4249/scholarpedia.42361

McCraty, R. (2017). New frontiers in heart rate variability and social coherence research: techniques, technologies, and implications for improving group dynamics and outcomes. *Front. Public Health* 5:267. doi: 10.3389/fpubh.2017.00267

Mitkidis, P., McGraw, J. J., Roepstorff, A., and Wallot, S. (2015). Building trust: heart rate synchrony and arousal during joint action increased by public goods game. *Physiol. Behav*. 149, 101–106. doi: 10.1016/j.physbeh.2015.05.033

Park, G., and Thayer, J. F. (2014). From the heart to the mind: cardiac vagal tone modulates top-down and bottom-up visual perception and attention to emotional stimuli. *Front. Psychol*. 5:278. doi: 10.3389/fpsyg.2014.00278

Park, H.-D., Correia, S., Ducorps, A., and Tallon-Baudry, C. (2014). Spontaneous fluctuations in neural responses to heartbeats predict visual detection. *Nat. Neurosci*. 17, 612–618. doi: 10.1038/nn.3671

Park, H.-D., and Tallon-Baudry, C. (2014). The neural subjective frame: from bodily signals to perceptual consciousness. *Philos. Trans. R. Soc. B* 369:20130208. doi: 10.1098/rstb.2013.0208

Rainville, P., Bechara, A., Naqvi, N., and Damasio, A. R. (2006). Basic emotions are associated with distinct patterns of cardiorespiratory activity. *Int. J. Psychophysiol*. 61, 5–18. doi: 10.1016/j.ijpsycho.2005.10.024

Rao, R. C. (1948). The utilization of multiple measurements in problems of biological classification. *J. R. Stat. Soc*. 10, 159–203. doi: 10.1111/j.2517-6161.1948.tb00008.x

Sato, N., Miyake, S., Akatsu, J., and Kumashiro, M. (1995). Power spectral analysis of heart rate variability in healthy young women during the normal menstrual cycle. *Psychosom. Med*. 57, 331–335. doi: 10.1097/00006842-199507000-00004

Sequeira, H., Hot, P., Silvert, L., and Delplanque, S. (2009). Electrical autonomic correlates of emotion. *Int. J. Psychol*. 71, 50–56. doi: 10.1016/j.ijpsycho.2008.07.009

Shaffer, F., and Ginsberg, J. (2017). An overview of heart rate variability metrics and norms. *Front. Public Health* 5:258. doi: 10.3389/fpubh.2017.00258

Shaffer, F., McCraty, R., and Zerr, C. L. (2014). A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. *Front. Psychol*. 5:1040. doi: 10.3389/fpsyg.2014.01040

Thayer, J. F., and Lane, R. D. (2000). A model of neurovisceral integration in emotion regulation and dysregulation. *J. Affect. Disord*. 61, 201–216. doi: 10.1016/S0165-0327(00)00338-4

Vallejo, M., Márquez, M., Borja-Aburto, V., Cárdenas, M., and Hermosillo, A. (2005). Age, body mass index, and menstrual cycle influence young women's heart rate variability. *Clin. Auton. Res*. 15, 292–298. doi: 10.1007/s10286-005-0272-9

Vesper, C., Abramova, E., Bütepage, J., Ciardo, F., Crossey, B., Effenberg, A., et al. (2017). Joint action: mental representations, shared information and general mechanisms for coordinating with others. *Front. Psychol*. 7:2039. doi: 10.3389/fpsyg.2016.02039

Vidaurre, C., Sander, T. H., and Schlögl, A. (2011). Biosig: the free and open source software library for biomedical signal processing. *Comput. Intell. Neurosci*. 2011:935364. doi: 10.1155/2011/935364

Yildirir, A., Kabakci, G., Akgul, E., Tokgozoglu, L., and Oto, A. (2001). Effects of menstrual cycle on cardiac autonomic innervation as assessed by heart rate variability. *Ann. Noninvas. Electrocardiol*. 7, 60–63. doi: 10.1111/j.1542-474X.2001.tb00140.x

Yoshino, K., Matsumoto, S., Someya, E., and Kitajima, M. (2011). Happiness and heart rate response: a case of fan services at Japanese professional baseball games. *Nat. Sci.* 3, 255–258. doi: 10.4236/ns.2011.33032

Zhang, J. (2007). Effect of age and sex on heart rate variability in healthy subjects. *J. Manipul. Physiol. Therap.* 30, 374–379. doi: 10.1016/j.jmpt.2007.04.001

Zhang, Y., Qin, F., Liu, B., Qi, X., Zhao, Y., and Zhang, D. (2018). Wearable neurophysiological recordings in middle-school classroom correlate with students' academic performance. *Front. Hum. Neurosci.* 12:457. doi: 10.3389/fnhum.2018.00457

Zhu, J., Ji, L., and Liu, C. (2019). Heart rate variability monitoring for emotion and disorders of emotion. *Physiol. Meas.* 40:064004. doi: 10.1088/1361-6579/ab1887

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

HYPOTHESIS AND THEORY
published: 14 August 2020
doi: 10.3389/fnhum.2020.00317

Coordination Dynamics: A Foundation for Understanding Social Behavior

Emmanuelle Tognoli[1,2], Mengsen Zhang[1,3], Armin Fuchs[1,4†], Christopher Beetle[4] and J. A. Scott Kelso[1,5]**

[1] Human Brain and Behavior Laboratory, Center for Complex Systems and Brain Sciences, Florida Atlantic University, Boca Raton, FL, United States, [2] Department of Biological Sciences, Florida Atlantic University, Boca Raton, FL, United States, [3] Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, United States, [4] Department of Physics, Florida Atlantic University, Boca Raton, FL, United States, [5] Intelligent Systems Research Centre, Ulster University, Londonderry, United Kingdom

Edited by:
Daniel Polani,
University of Hertfordshire,
United Kingdom

Reviewed by:
Hossein Ehsani,
University of Maryland, United States
Alexander Maye,
University Medical Center
Hamburg-Eppendorf, Germany

***Correspondence:**
Emmanuelle Tognoli
tognoli@ccs.fau.edu
J. A. Scott Kelso
jkelso@fau.edu

†Deceased

Specialty section:
This article was submitted to
Motor Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

Received: *04 February 2020*
Accepted: *17 July 2020*
Published: *14 August 2020*

Citation:
Tognoli E, Zhang M, Fuchs A,
Beetle C and Kelso JAS (2020)
Coordination Dynamics: A Foundation
for Understanding Social Behavior.
Front. Hum. Neurosci. 14:317.
doi: 10.3389/fnhum.2020.00317

Humans' interactions with each other or with socially competent machines exhibit lawful coordination patterns at multiple levels of description. According to Coordination Dynamics, such laws specify the flow of coordination states produced by functional synergies of elements (e.g., cells, body parts, brain areas, people...) that are temporarily organized as single, coherent units. These coordinative structures or synergies may be mathematically characterized as informationally coupled self-organizing dynamical systems (Coordination Dynamics). In this paper, we start from a simple foundation, an elemental model system for social interactions, whose behavior has been captured in the Haken-Kelso-Bunz (HKB) model. We follow a tried and tested scientific method that tightly interweaves experimental neurobehavioral studies and mathematical models. We use this method to further develop a body of empirical research that advances the theory toward more generalized forms. In concordance with this interdisciplinary spirit, the present paper is written both as an overview of relevant advances and as an introduction to its mathematical underpinnings. We demonstrate HKB's evolution in the context of social coordination along several directions, with its applicability growing to increasingly complex scenarios. In particular, we show that accommodating for symmetry breaking in intrinsic dynamics and coupling, multiscale generalization and adaptation are principal evolutions. We conclude that a general framework for social coordination dynamics is on the horizon, in which models support experiments with hypothesis generation and mechanistic insights.

Keywords: HMI, HRI, Coordination Dynamics, social interaction, metastability, multiscale, complex systems, Human Dynamic Clamp

INTRODUCTION

Social systems nest very small structures, the molecular, genetic and cellular machinery of living things, into progressively larger structures – all the way up to entire organisms engaged in mutual interaction with the environment and with each other. Quite crucially and across all levels, the parts (e.g., organelles, organs, organisms, organizations) coordinate dynamically with other parts, engaging and disengaging within and between their respective coalitions and across levels (upward∼downward causation, e.g., genes or neurons influencing social behavior, and vice-versa). A main goal of our research program is to find general systems of equations – expressing lawful regularities – that explain social systems' coordination dynamics within and across levels, irrespective of level-specific details (Oullier and Kelso, 2009; see also Kelso, 2009a; Kelso et al., 2013). We approach this goal by examining dynamic coordination patterns empirically; embedding those observations in mathematical models; and returning to empirical data to verify newly arisen predictions, in a recursive manner. For present purposes, a model is the recreation of a natural system's key behavior that facilitates its understanding. "Understanding" is sought, not through some privileged scale of analysis, but within the abstract level of the essential collective variables and their coordination dynamics, regardless of scale or material substrate (Kelso et al., 1987; Schöner and Kelso, 1988b).

Quite a few modeling frameworks have been applied to social systems, including agent-based models (e.g., Axelrod, 1997; Gilbert and Terna, 2000; Bonabeau, 2002; Schweitzer, 2007), cellular automata (e.g., Hegselmann, 1998; Batty, 2007), Lotka-Volterra (e.g., Matsuda et al., 1992; Castellano et al., 2009), stochastic diffusion (e.g., Arató, 2003; López-Pintado, 2008; Kimura et al., 2010), Bayesian (e.g., Yang et al., 2011), Markov (e.g., Singer and Spilerman, 1976; Gintis, 2013), signal flow graphs and block diagrams (Liu and Ma, 2018), recurrent networks (Irsoy and Cardie, 2014), and, central to this review, the HKB model (after Haken et al., 1985; see also Schöner and Kelso, 1988b; Kelso, 1995; Tognoli et al., 2018a), itself based on the concepts of synergetics (Haken, 1977) and the mathematical tools of non-linearly coupled non-linear oscillators. One of the key strengths of the HKB model (and its numerous extensions) is its possession of *intrinsic dynamics* (Kelso, 1995, Ch 6). That is, the system of equations is formalized from two sides: one supplying the intrinsic dynamics of the unit (what it does when left alone to itself), and the other – whose significance social scientists will recognize – reflecting constraints imposed by relation(s) with other units. Intrinsic dispositions and social influences are complementary aspects of social interaction, without which an agent would be carried along by the ebbs and flows of whatever jolts it encounters (see also Friston, 2011 and Kostrubiec et al., 2012 for related views).

Mathematical models, when combined with theoretical concepts, have the power to accomplish an important aim for research that aspires to characterize cross-scale relations: they permit widely different phenomena to fall under common scrutiny. Numerous examples in the history of science speak to the colossal payoff that follows successful integration across scales. Newton's famous unification of the laws that govern the fall of the apple and the motion of celestial bodies comes to mind. Deterministic chaos and quantum mechanics contain many more examples, although it must be said that "emergent phenomena" exist as well, i.e., where the whole is not only greater than the sum of its parts, but different too (Anderson, 1972; Haken, 1977; Laughlin and Pines, 2000).

How then is one to approach the daunting diversity of dynamical behaviors that is encountered across scales of observation? Our paradigm involves collective variables and non-linear oscillators on the mathematical side, rhythmic finger movements as basic observational units at the behavioral level (Kelso, 1981, 1984; Haken et al., 1985), and of course, neural oscillations at the neural level (see Schöner and Kelso, 1988b; Glass, 2001; Tognoli and Kelso, 2014b). Oscillations might be considered as the "ground zero" of open dynamical systems, their most elemental form. First, temporal symmetries of limit cycles (the mathematical structure for oscillations) allow for fruitful mathematical simplifications. Second, oscillations are pervasive in nature and obvious in their simplest form at the inception of complex organisms: spontaneous oscillations are found in neural and motor activity prenatally (Robertson, 1993; Khazipov and Luhmann, 2006) and destined to endure, albeit in more complex form, throughout the life of living systems (Turrigiano et al., 1994; Gal et al., 2010; Marom, 2010). Evidence for the primordial role of oscillations in human behavior also comes with the occurrence (possibly unmasking) of repetitive movements in several developmental and aging disorders (Brown, 2003; Abbott et al., 2017). There are clear signs that oscillations are exploited for subcortical control (Taga et al., 1991; Grillner et al., 1998; Stewart, 1999; Righetti et al., 2005) and similar hypotheses have been proposed for the cortical level (Edelman and Mountcastle, 1978; Kelso and Tuller, 1984; Yuste et al., 2005; Buzsáki, 2010; Tognoli and Kelso, 2014b). The primacy of oscillations for the regulation and control of living systems has been articulated in the early works of Iberall, Yates, Morowitz, and others. For example, Homeokinetic theory (e.g., Soodak and Iberall, 1978; Yates, 1982) addresses the conditions for persistence, autonomy and self-organization in biological systems from a physical perspective (irreversible thermodynamics). A fundamental tenet is that energy flow from a source to a sink will lead to at least one cycle in the system (Morowitz, 1968). In the homeokinetic view, control is effected by means of coupled ensembles of limit cycle oscillatory processes. Limit cycle oscillations represent the only temporal stability for non-conservative, non-linear systems, that is, they are capable of making up for naturally occurring dissipative losses. Loose coupling of limit-cycle processes exists at all scales. Among their attractive features are their self-sustaining properties, their ability to operate independently of initial conditions, their stability in the face of moderate perturbations, and, perhaps most important the properties of mutual entrainment and synchronization (Minorsky, 1962; Kelso et al., 1981; Winfree, 2001). Furthermore, elaboration of arbitrarily more complex dynamics can be obtained from oscillatory functions, as suggested by the work of pioneer mathematicians like Joseph Fourier, showing some bridges between oscillations and ordinarily irregular dynamics. On both physical and mathematical grounds, therefore, it follows that a path from simple oscillatory dynamics to more complex dynamical behavior (typically observed across multiple scales in biological systems) may be possible.

In the following, we will present briefly our protracted efforts to build more complex dynamics from the paradigm initiated by Kelso and colleagues (see also Avitabile et al., 2016). Before that, we describe briefly the experimental paradigm, analysis strategy, theory and models that constitute the mathematical building blocks of our approach to understanding social coordination. The idea is to show how we grow our modeling paradigm toward increasingly more life-like situations and to identify where the edges of future advances may lie.

EXPERIMENTAL PARADIGM

A minimal experimental paradigm for social coordination dynamics has two people exchanging information by virtue of their senses and effectors as shown in **Figure 1**, forming a figure eight that has maximal symmetry. In our canonical experiments (Tognoli et al., 2007; Oullier et al., 2008; Tognoli, 2008; Tognoli and Kelso, 2015), action is generated by the index finger (often depicted by a phase angle of the finger relative to the joint) and perception is mainly through vision (though both details are certainly amenable to other arrangements within the sensorimotor repertoire of our human subjects, see e.g., Schmidt et al., 1990). This choice does not ignore the fact that plenty of social behaviors (conversations, emotions, dance, jamming, etc.), at first seem to carry much greater sociocultural significance. In the pure tradition of physics, stripping away complications is a strategy to draw out simple controllable pieces of the system and set them in motion in ways that are tractable for mathematical models and their experimental counterparts (see also Kelso, 1995; Stewart, 1999; Nowak, 2004). Our work seeks laws of coordination (subject to proper empirical verifications whenever feasible) that transcend the specific choice of effectors (sweat gland, vocal tract, facial, limb muscles etc.) or sensory pathways in order to draw general mathematical foundations. We start with the simplest dynamics that can be experimentally manipulated and understood theoretically.

With the simple paradigm of two people moving their fingers back and forth in view of one another, we are able to obtain continuous state variables describing the trajectory of each participant's effector at the behavioral level, their coordination dynamics (viz. the relative phase between the two finger movements, see "Order parameter" below; Tognoli et al., 2007; Oullier et al., 2008; Tognoli, 2008; see also Schmidt et al., 1990, 2011; Richardson et al., 2007; Schmidt and Richardson, 2008; Marsh et al., 2009; Janata et al., 2012; Reddish et al., 2013; Fine and Amazeen, 2014; Fusaroli et al., 2014; Keller et al., 2014; Tschacher et al., 2014; de Poel, 2016; Moreau et al., 2016; for a variety of related approaches). Further, when expanding this work, information can be gained about concomitant activities in the brain (Tognoli et al., 2007; Jantzen et al., 2008; Naeem et al., 2012; see also Hari and Kujala, 2009; Dumas et al., 2011; Sänger et al., 2011; Konvalinka and Roepstorff, 2012; Pfeiffer et al., 2013; Babiloni and Astolfi, 2014; Cacioppo et al., 2014; Hirata et al., 2014; D'Ausilio et al., 2015; Koike et al., 2015; Zhou et al., 2016; Kawasaki et al., 2018; Mu et al., 2018; Pezzulo et al., 2019), and in emotional subsystems (e.g., Zhang et al., 2016; see also Anders et al., 2011; Balconi and Vanutelli, 2017; Reindl et al., 2018). Objective measures of brain and behavior offer tight systems of constraint that connect experiments and experimentally validated models; their continuous nature serves well a modeling framework that uses collective variables/order parameters at the coordinative level and coupled oscillators as components. Finally, such collections of neurobehavioral oscillations aptly embed the multiscale and reciprocal nature of self-organizing processes that play out in social systems (Coey et al., 2012).

ORDER PARAMETER: CONNECTING MODELS AND EXPERIMENTS

A key concept of Coordination Dynamics – following along the lines of Synergetics (Haken, 1977) – is the collective variable or order parameter which has been demonstrated to cut across different kinds of parts and processes (and across levels) thereby dissolving traditional (and somewhat arbitrary) divisions (e.g., between "cognitive" and "motor") and enabling a novel, dynamic framework for understanding collective/social behavior (Kelso, 1995, 2009a,b, 2012; Tognoli, 2008; Coey et al., 2012; Richardson et al., 2014; Tognoli et al., 2018a; see also

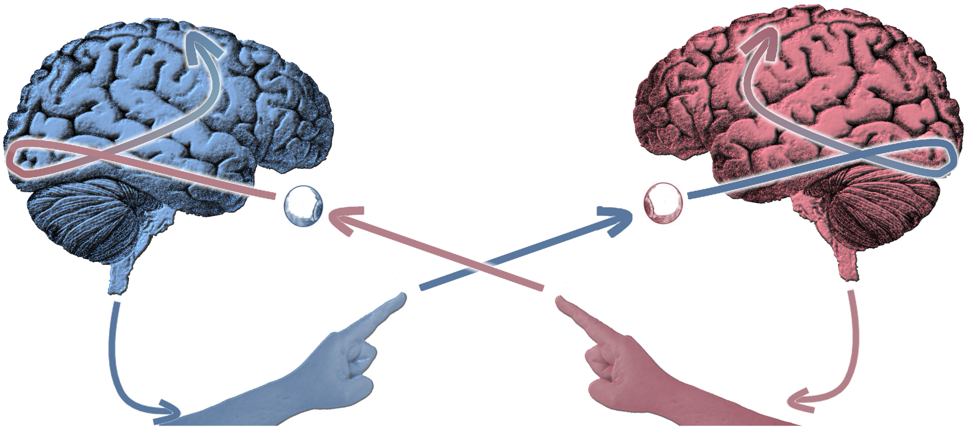


FIGURE 1 | Schematic of the experimental paradigm of social coordination dynamics in which two participants simultaneously perceive and produce behavior in view of each other. More specifically, in our canonical experiment, subjects move their fingers in continuous fashion while at the same time observing their partner doing the same. The paradigm's simultaneity of dyadic perception and action – bidirectional coupling – is geared toward observing self-organizing processes.

Macrae and Miles, 2012). In Coordination Dynamics, the order parameter tracks the relation between parts of a system in time (when they come to work together and when they go apart) – that is, the dynamics of their *coordination* in contrast to the dynamics of state variables characterizing the individual component parts. We therefore resort to the relative phase (the difference ϕ between the phase of each oscillator ϕ_1 and ϕ_2), as the collective variable specified in the equations (**Figure 2A**) and empirically scrutinized in experimental and modeling data (Fuchs and Kelso, 1994, 2009, 2018; Kelso, 1995, 2009a; Lagarde et al., 2006; Kelso and Tognoli, 2007; Tognoli and Kelso, 2009; Fuchs, 2014). From the standpoint of component rhythmic behavior, the HKB model was built from phase-coupled oscillators and accordingly, the theory indicates synchrony or synchronous tendencies as target phenomena (Haken, 2013; see also Strogatz and Stewart, 1993; Bennett et al., 2002; Pikovsky et al., 2002).

Starting in the limit case of strongly coupled systems such as two entrained gears that are so perfectly fit to each other that there isn't room for either phase trajectory to deviate from the other, the resulting relative phase is constant (**Figure 2B**, yellow). It is obvious that such a strong and immutable coupling cannot harbor the adaptation that is the hallmark of intelligent living systems, including instabilities, phase transitions and metastability (Kelso, 1995, 2001, 2010, 2012). Complex systems having the necessary propensity for adaptation and reorganization especially entail the latter feature (**Figure 2B**, magenta, green), that is, parts and coalitions of parts alternate between cooperative tendencies (parts transiently binding, manifested by transiently horizontal epochs, or dwells, of the relative phase) and release from them (phase wrapping; Kelso, 1991). In the other limit when coupling vanishes to zero, the resulting uncoupled system typically manifests a flat diagonal trajectory of its relative phase, reflecting the mere incidence of their respective intrinsic phase dynamics (**Figure 2B**, blue). Stable trajectories (**Figure 2B**, yellow) are found in regimes with attractors (**Figure 2A**, for small values of $\delta\omega$, see right bracket); metastable trajectories in regimes sans attractors (**Figure 2A**, for larger values of $\delta\omega$ past a critical threshold, see left bracket; and weak coupling, not shown).

CLASSICAL EXPERIMENTAL FINDINGS ON SOCIAL COORDINATION

Under the simple paradigm presented in section "Experimental Paradigm" and with the analysis tools of section "Order Parameter: Connecting Models and Experiments," we review a number of experimental results supporting the claim that coordination dynamics at multiple scales is metastable, i.e., intermingled dwells and escape in the relative phase exist (relative phase trajectories in **Figure 3**). We also found what appears to be persistently stable relative phases (not shown, see Tognoli et al., 2007; Kelso et al., 2009; Dumas et al., 2014 for examples), which either pertain to genuine phase locking (dynamical regime *with attractors*) or metastable dwells (*sans-attractor*) whose characteristic time scale exceeds the window size. Findings of metastability span dyadic social coordination (Tognoli et al., 2007; Oullier et al., 2008; Tognoli, 2008; **Figure 3A**); experiments with a "Human Dynamic Clamp" in which one member of the dyad is a computational model of a surrogate social partner governed by the HKB equations (**Figure 3B**, Kelso et al., 2009); multiagent social coordination (**Figure 3C**; Zhang et al., 2018, 2020); and, outside our paradigm but hinting at the phenomenon's generality, the collective flashing of fireflies (**Figure 3D**; Tognoli et al., 2018b) that we found not to be fully synchronized *stricto sensu* (recalling that synchronization requires attractors). As predicted from theory, when parts have extensive symmetries in their intrinsic dynamics (small $\delta\omega$, **Figure 2A**), the attracting tendencies (**Figures 3A–C**: histograms of the relative phase) have concentrations at inphase (both oscillator trajectories rise and then fall in step) and antiphase (one oscillator rises when the other falls), befitting the extended HKB model as a general model of behavioral and neural coordination (Kelso, 1991; Kelso and Haken, 1995). These observations are also consistent with HKB model predictions (Kelso et al., 1987; Kelso, 2008). For instance, coordination can be weakened by faster dynamics (Kelso et al., 2009), weaker coupling imposed by one participant (as in the "parametrizable" human dynamic clamp: Kelso et al., 2009), or greater diversity of prior dispositions (Zhang et al., 2018, 2020).

In a parallel body of work, we examined the coordination dynamics of brain activity upon which social behavior is predicated, in particular in the domain of electrophysiology. Here again and at several scales, we found evidence of synchronization tendencies or metastability (Tognoli and Kelso, 2009, 2014b). The hypothesis that synchrony (as distinct from synchronization *tendencies*, see below) underlies the binding of local oscillatory processes has taken root in neuroscience (e.g., Gray, 1994; von der Malsburg, 1995; Singer, 1999; Bressler and Kelso, 2001; Varela et al., 2001; Bressler and Tognoli, 2006; Uhlhaas et al., 2009; Buzsáki, 2010; van Wijk et al., 2012; Harris and Gordon, 2015), with the caveat that due to the inherently dipolar nature of electromagnetic fields, electrophysiological data at all scales are replete with spurious inphase and antiphase coordination (Freeman, 1980; Nunez et al., 1997; Nolte et al., 2004; Pascual-Marqui, 2007; Tognoli and Kelso, 2009; Van de Steen et al., 2019). Therefore, empirical evidence needs to be heeded carefully before rendering definitive interpretations of remote synchrony between neural ensembles. Initially setting aside inphase and antiphase (due to the fact that they contain both true and spurious synchronization), we were able to discover transient synchronization patterns (see, e.g., **Figure 4** from Tognoli and Kelso, 2009). In the context of our empirical quest to discriminate stability from metastability, synchrony was suggestive of stable states and state transitions–thereby seeming to point to attractors and bifurcations as governing brain dynamics. We however found that a discontinuous spatiotemporal organization was not definitive proof of attractor-based states and transitions in brain dynamics. It is also entirely compatible with the alignment and misalignment of phases across scales that is the hallmark of metastability (e.g., see **Figure 5** from Tognoli and Kelso, 2014b). Moreover, relative phase trajectories, with faint dwells and limited periods of common frequencies, point toward weak coupling. The latter is consistent with the idea that human

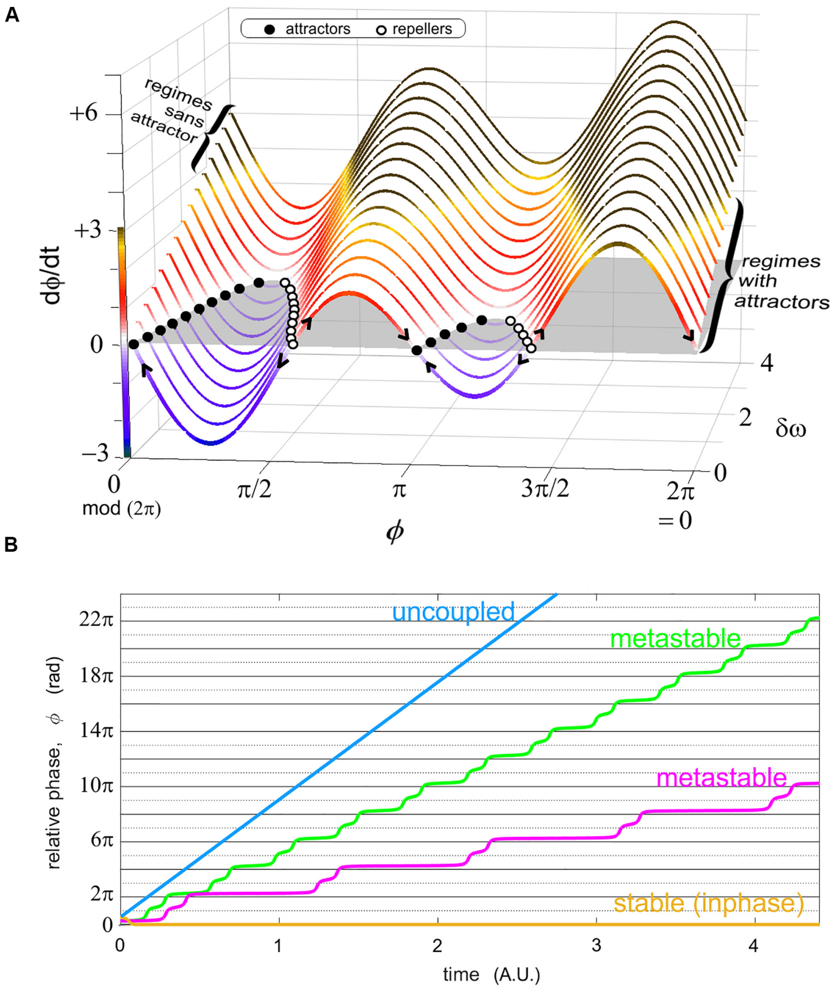


FIGURE 2 | Two representations of the relative phase as an order parameter connecting models and experiments. **(A)** shows the phase portrait of ϕ in the "extended" HKB model (Kelso et al., 1990) for various values of a diversity parameter δω. This graph carries regimes of coordination with attractors in the front of the figure (for modest diversity δω, shown; also when coupling is strong, not shown) and those without attractors (large diversity of the parts, shown in the back of the figure; and/or weak coupling, not shown). Attractors exist when the phase portrait (a function describing the rate of change of the relative phase as a function of itself) has values at $\frac{d\phi}{dt} = 0$ (i.e., the coordination does not change over time) and a converging flow (filled black dots attracting the flow as indicated by the arrows). Stable regimes reflect a sustained cooperation among the system's parts, but this stability also leads to inflexibility (see Kelso and Tognoli, 2007 for more details). In **(B)**, four sample (unwrapped) relative phase evolutions over time illustrate stable coordination (yellow) where the order parameter ϕ persists at the same value (ad infinitum in models); metastability (magenta, green) with their characteristic dwells (quasi horizontal epochs, attracting tendencies near inphase, i.e., 0 rad. and antiphase, i.e., π rad. modulo 2π) and escape (wrapping); and uncoupled behavior (blue), whose relative phase grows continually with time (it approaches a linear function when the probability distribution of individual phases lacks remarkable joint phase ratios).

brains are subject to the coordinative demands of many local populations that trade among each other due to ubiquitous weak coupling. Definitive evidence in favor of either dynamics with or *sans* attractors remains tentative, and its resolution likely resides in perturbation approaches to brain dynamics. Furthermore, because of the pervasive ambiguity of spurious inphase and antiphase in the EEG and the evasiveness of non-inphase dwells from lower scales (Tognoli and Kelso, 2009), it remains difficult to offer definitive evidence of bistable tendencies in brain electrophysiological patterns, and test this specific prediction from the HKB model. Phase transitions, however, have been established and have suggested bistable tendencies in MEG (Kelso et al., 1991, 1992; Fuchs et al., 1992, 2000b) and fMRI dynamics, starting with Meyer-Lindenberg et al. (2002) (see also Fox et al., 2005; Tognoli and Kelso, 2014b).

A CHRONOLOGY OF THE HKB MODEL'S AUGMENTATIONS

The HKB system of equations was initially developed to model a non-linear, self-organizing phenomenon discovered in human movement coordination (Kelso, 1981, 1984). When simultaneously set in motion, two homologous body parts, for instance left and right hand, may be stabilized in either of two patterns, inphase or antiphase (suggestive of attractors).

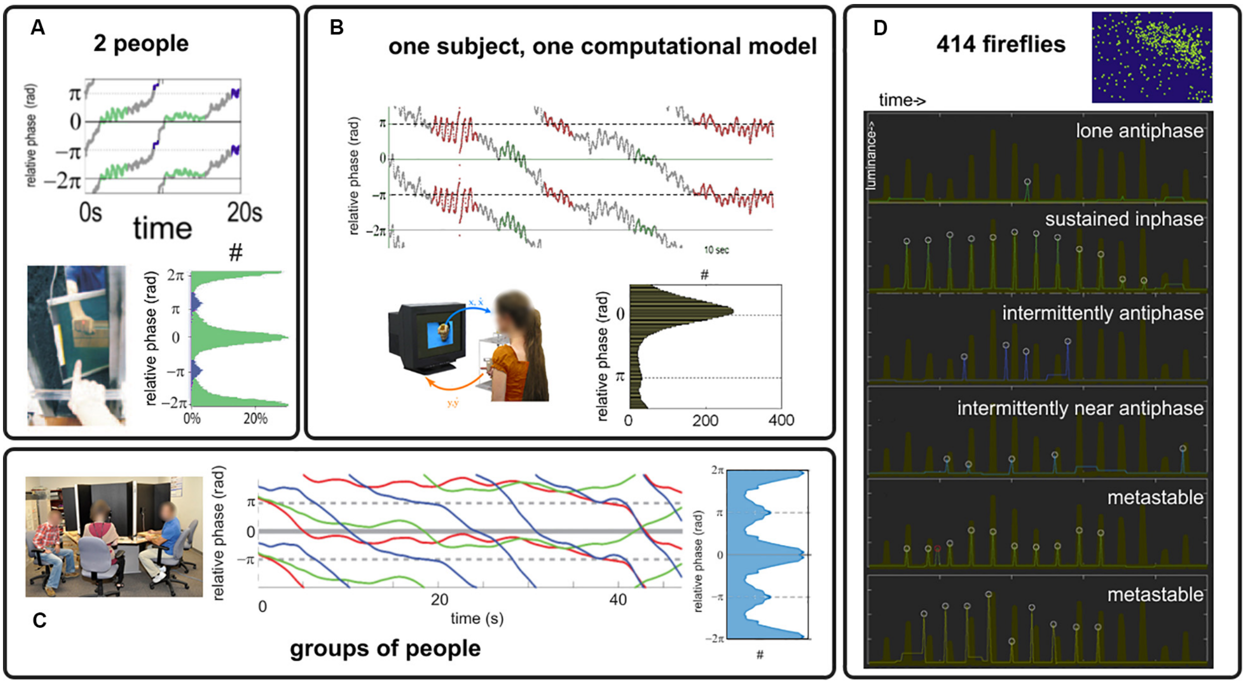


FIGURE 3 | Behavioral coordination in human dyads **(A)**, human interaction with a Human Dynamic Clamp **(B)**, ensemble of people **(C)** and to hint at generality, in fireflies flashing [white circles, **(D)** the thick brown trace indicates the collective behavior of their large colony]. Relative phase trajectories from **(A–C)** are chosen to illustrate an alternation of dwells and escapes that is typical of metastability. Attracting tendencies are found toward inphase and antiphase [see histograms in **(A–C)**, with count identified with the symbol #, showing relative phase distribution for the most strongly coupled conditions in their respective paradigms, see original references for details and for examples of weaker organization].

As movement frequency was increased, however, there was an abrupt switch wherein one pattern lost stability and gave way to the other (here antiphase to inphase). In dynamical system parlance, this switching behavior can be described as an order-to-order phase transition, corresponding mathematically to a bifurcation. The HKB model was built to capture bistability, bifurcation and hysteresis (the finding of different critical frequencies when approaching the bifurcation with decreasing or increasing movement frequencies). Those three elements are characteristic of many complex non-linear systems. Within a multiscale paradigm, it is natural to ask whether governing laws retain their validity at proximal scales (Gell-Mann, 1988; Jirsa and Haken, 1997; Kelso et al., 1999). Our group started to ask whether the HKB equations governing the coordination between two hands of the same person applied when we scale up one level: two hands, one each from two different individuals. Thus, we forayed into social coordination dynamics (Tognoli et al., 2007, 2018a,b; Oullier et al., 2008; Tognoli, 2008; Kelso et al., 2009, 2013; Kelso, 2012; Dumas et al., 2014, 2018; Kostrubiec et al., 2015; Tognoli and Kelso, 2015; Nordham et al., 2018; Zhang et al., 2018, 2019). Social coordination dynamics complemented the same question previously asked at the level of the individual: the hand-owners' neuro-muscular system studied with neurophysiological tools, MEG, fMRI, and EEG (e.g., Tuller and Kelso, 1989; Kelso et al., 1992, 1998; Wallenstein et al., 1995; Jirsa et al., 1998; Mayville et al., 1999, 2002; Fuchs et al., 2000a,b; Jantzen et al., 2004, 2009; Oullier et al., 2004; Jantzen and Kelso, 2007; Kelso and Tognoli, 2007; Tognoli and Kelso, 2009, 2014b, 2015; De Luca et al., 2010; Banerjee et al., 2012; and many others). Since a sizable body of experimental work had supported HKB's relevance both in theoretically predicted and *de novo* phenomena (section "Classical Experimental Findings on Social Coordination"), we have adopted HKB as the root model in our social neuroscience research program.

Above, we justified the use of simple behavioral paradigms that were the key to HKB's initial development. In the following, we demonstrate several advances toward further generalization in the context of social coordination dynamics. Though by no means inclusive[1], **Figure 4** presents six stages of the HKB framework. The original equations (**Figure 4**, eq. 1; Haken et al., 1985) describe (at the oscillator level on the left and collective level on the right) two non-linearly coupled non-linear oscillators x and y with intrinsic frequency Ω (eigenfrequency), that are bound by a coupling function with notable coupling strength constants a and b. The existence of both coupling terms, one with period ϕ and the second with period 2ϕ, are responsible for the bistability of both inphase and antiphase, and the b/a

---

[1] For example, other developments not considered here include learning, recruitment of new degrees of freedom, parametric stabilization, multifrequency coordination, trajectory formation, stabilization of unstable systems, intentional switching, etc. Most of these aspects have been developed within single subject situations and have yet to be systematically explored in social contexts, our focus here. They do, quite obviously, provide a reservoir for future work in social coordination dynamics.

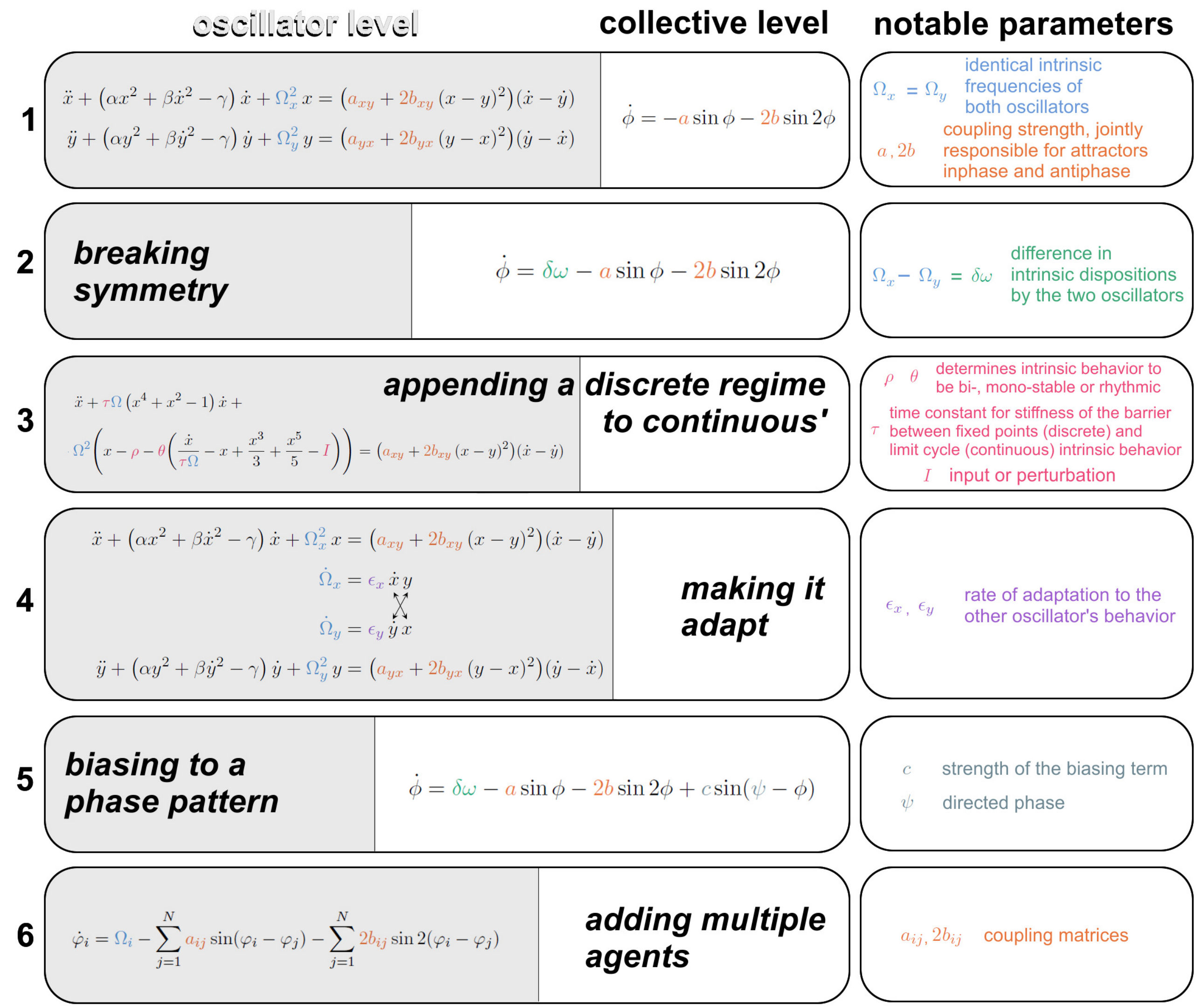


FIGURE 4 | Augmentations of the HKB model toward increasing life-like relevance. **(1)** The first equation is spelled out at the oscillator level on the left (specifying the position, velocity and acceleration of oscillators x and y, with notable parameter Ω carrying the intrinsic frequency, see "Introduction" for motivation); and at a collective level on the right (specifying the rate of change of x and y's relative phase ϕ. The coupling strength parameters a and b are responsible for the bistability of inphase and antiphase. **(2)** A symmetry breaking term δω (Kelso et al., 1990) gives rise to metastability. **(3)** A modification to the oscillators' dynamics supplies regimes with discrete behavior -excursion from rest- and continuous cyclical behavior in a single formalism (Jirsa and Kelso, 2005), whose topology is controlled by three parameters ρ, θ, and τ. Transitions between regimes are autonomously followed in a human-machine interaction (Dumas et al., 2014). **(4)** The intrinsic dynamics of the oscillators Ω is coupled with adaptation rate ϵ for partners to adapt their more stable behavioral dispositions as they interact (Nordham et al., 2018). **(5)** A bias parameter Ψ is introduced to direct the coordination to pull human-machine dyads away from their spontaneous attractors with strength c (Dumas et al., 2014). **(6)** A scalable multiagent system of equations is built upon empirical data (Zhang et al., 2019) that carries the first and second order coupling term just like the original HKB model from eq. 1, under coupling matrices a and b. See references in text.

ratio controls the "depth" of each attractor basin. In the case of eq. 1, $\Omega_x = \Omega_y$: the two oscillators have identical intrinsic dynamics, a sensible approximation given the model's origin with two homologous body parts.

In nature and especially in complex biological and social systems, it is seldom the case that coordination exclusively concerns perfectly matching pieces of machinery: many components ought to work across the divide of different intrinsic dispositions (a child and a father pacing together on a beach; brain areas generating beta and mu oscillations…). The second equation from **Figure 4**, eq. 2; Kelso et al. (1990) breaks symmetry in the intrinsic behavior of the oscillators, letting Ω_x differ from Ω_y. At the collective level (Kelso et al., 1990; for equations at the component level see Fuchs et al., 1996), this extended HKB model results in a new term δω whose existence has two main effects: it shifts the attractors away from inphase and antiphase coordination and it shrinks the basin of attraction until the weaker antiphase, followed by the stronger inphase, eventually vanish, thereby unlocking a metastable regime that retains attracting tendencies (effective pooling of collective effort), but has lost its fixed points *stricto sensu* (see also **Figure 2A**). Such metastable regimes are important for their enhanced agility to continually disband coordination patterns and create new ones (Kelso, 1991, 1995; Kelso and Tognoli, 2007).

Obviously from common experience, at some level, functionally adapted systems do not continuously spin the wheel of their dynamics ceaselessly. Numerous processes retain cyclical characteristics under entrainment with the environment or even intrinsically (Glass and MacKey, 1988; Strogatz and Stewart, 1993; Winfree, 2001; Tognoli and Kelso, 2014b; and references in Introduction). But rest is also a characteristic biological behavior; and goal-directed agents, for instance people or brain networks (Attwell and Laughlin, 2001; Kramer and McLaughlin, 2001), tend to set in motion only briefly under intent, perhaps to manage energy constraints (Attwell and Laughlin, 2001). This propensity to switch from cyclical to resting behavior was built into the Excitator model (**Figure 4**, eq. 3, from Dumas et al., 2014, modified after Jirsa and Kelso, 2005), inspired from the explicitly discrete nature of neuronal excitability (and human movement coordination, Kelso et al., 1979). Its right-hand side (the coupling term) is identical to eqs. 1–2 but its left-hand side has terms that, under parametric control of a separatrix, part the flows of the coordination dynamics into discrete and continuous domains. HKB's augmentation to discrete dynamics is a building block for many diachronic social behaviors such as turn-taking in talks or mere conspecific observation (e.g., Dumas et al., 2010, 2014; Kawasaki et al., 2013; Tognoli and Kelso, 2015; Pérez et al., 2017) and is intimately linked to delayed coordination and the emergence of roles amongst social participants (Dumas et al., 2014).

Thus far, individual characteristic behavior in isolation has been left unchanged by interaction with others. Now, it is well-recognized in sociology and neuroscience that parts (e.g., people, brain areas) do in fact change (see also Newell et al., 2008), and specifically according to their history of interaction with other parts. This is a truism of adaptation. Taking inspiration from Righetti et al. (2006)'s work on frequency learning in oscillators, Dumas et al. (2014) created an adaptive "Human Dynamic Clamp" based on earlier research on Virtual Partner Interaction (Kelso et al., 2009). In this work, Dumas et al. (2014) used a human to broaden the HKB model's behavioral repertoire: by making its intrinsic frequency adapt to a human partner's input, a virtual partner was conditioned to track a slowly evolving range of movements that a human would produce in its "view." This work opened the way to social adaptation in our mathematical models. It was the first incursion into broader behavioral repertoires for HKB-based surrogate social partners, who grew new capabilities by virtue of their interaction with humans.

In the Dumas et al. (2014) work, however, a relaxation term returns the oscillator to its characteristic frequency preference. Adaptation has one more sweeping phenomenology to hand over to the model and it concerns what happens after, not during the interaction. In the paradigm of dyadic social coordination, Oullier et al. (2008) discerned an increased spectral overlap after episodes of social coordination ("social memory"). Additionally, Tognoli (2008) found the persistence of stable relative phase past the period when visual information exchange secures the binding of both oscillatory trajectories. This aftereffect was further studied by Nordham et al. (2018), who teased apart three factors modulating the strength of social memory: coordinative stability, coupling strength (stronger aftereffects for more stable and more strongly coupled trials) and initial frequency differences (stronger aftereffects in trials with smaller initial differences). It turned out that all three factors arise from a common source. In the modeling section of the same paper, Nordham et al. (2018) showed that a universe of experimental observations in interactional behavior (e.g., strong aftereffect in both partners, in one or in none; multiple precursor conditions before and during interaction that influenced behavior post-interaction) could be accommodated by a single modification of the model. Each oscillator's intrinsic frequency ceased to be a constant. Instead, it became an equation that conjugated self's and partner's dynamics, weighted by an idiosyncratic parameter ϵ – the individual propensity to let one's self be attracted (or sometimes repelled) by the other (**Figure 4**, eq. 4). In this adaptive HKB model, social memory deploys all of its forms out of a combination of three parameters: frequency adaptation, coupling strength and initial frequency difference.

Now let's consider the phase pattern at which social coordination happens. While experimental data demonstrate that effortless coordination between two people occurs at or near inphase and antiphase (Section "Classical Experimental Findings on Social Coordination" and Schmidt et al., 1990; Richardson et al., 2007; Marsh et al., 2009; Fine et al., 2015; though see Avitabile et al., 2016 for nuanced theoretical and empirical insights), there is plenty of skillful interpersonal coordination that ought to happen against the grain of "natural" tendencies, e.g., in orchestra playing (Walton et al., 2015) or in the performance of skillful joint actions (Duarte et al., 2012; Issartel et al., 2017). An experimental line of research on learning had shown the remodeling of attractive states/tendencies as they are subjected to practice (Schöner and Kelso, 1988b; Zanone and Kelso, 1992, 1997; Kelso and Zanone, 2002; Kostrubiec et al., 2012, 2015). Change in the attractor landscape was modeled by Schöner and Kelso (1988a) as a task requirement ("informational forcing") that tuned the locus of the attractors. Considering that conspecifics are a crucial part of the environment (MacMahon et al., 1978; Dunbar and Shultz, 2007; Barsalou, 2008; Adolphs, 2009), Dumas et al. (2014) expanded the additional forcing term from Schöner and Kelso (1988a) into a Virtual Partner parameter to lure humans into collective behaviors that would *a priori* be unstable for human dyads to perform. To do so, the Virtual Partner's mathematical model was augmented with a biasing term that attracted the collective dynamics to a target phase Ψ with strength C (**Figure 4**, eq. 5). Incorporating this feature into the Human Dynamic Clamp, Kostrubiec et al. (2015) demonstrated that spontaneously unstable patterns of phase coordination (e.g., at 90°) could be coproduced by a human and a Virtual Partner, the latter set with a strong bias to teach that pattern to the human. Some degree of learning was also corroborated, and the system was nicknamed "Virtual Teacher." This model augmentation ventures into social contexts where the participants' roles are markedly different: a computationally forged bias (representing the timescale of attracting structures) allows transfer of coordinative patterns between participants, characteristic of social learning.

All the models exposed above pertain to dyadic situations. Early on in Coordination Dynamics, the question of multiagent models had been addressed by Schöner et al. (1990) and Jeka et al. (1993) in the context of "quadrupedal" coordination patterns of people's upper and lower limbs in which the (collective) state variables of the system constitute 3 relative phases. But proposals to scalable n-dimensional HKB systems remained to be achieved. Zhang et al. (2018) reasoned that useful foundations rested with empirical data at the intermediate scale (somewhat bigger than two to make room for groupings within groupings, but smaller than the experimentally unattainable infinity; and importantly, tractable for a detailed analysis of phase coordination patterns). To constrain the model with empirical data (and with the underlying goal of integrating across levels, see "Introduction"), Zhang et al. (2018) set up the perceptual-motor coupling of eight people whose movements caused taps on a touchpad and who saw everyone's taps as flickering patterns on spatially situated LED arrays (**Figure 3C**). The subsequent model development explored a variety of frameworks before settling on a hybrid system of HKB and Kuramoto equations (Zhang et al., 2019; **Figure 4**, eq. 6; including a coupling term with connectivity matrix a_{ij} fully equivalent to the Kuramoto model and another term with connectivity matrix b_{ij} resembling the second order coupling term from HKB, see eqs. 1,2,5). Development of the model built on useful functional symmetries carefully crafted in the experimental setup. The combination of extended HKB and Kuramoto fulfilled all the empirical constraints provided by Zhang et al. (2018) across multiple levels of description. Importantly, the Kuramoto model alone was insufficient because it did not uphold co-expressed inphase and antiphase patterns that are central features of HKB and that the eight agent experiment had uncovered. With this scalable system now made relevant by experimental data on human multiagent coordination, an opportunity exists to relate coordination dynamics across multiple scales, and especially across levels of description spanning the neural, behavioral and social [for an interesting related approach that also uses HKB coupling, see the 'sheep herding' paradigm of Nalepka and colleagues (Nalepka et al., 2017, 2019)].

CONCLUSION AND OUTLOOK

To complement many approaches that focus on a unique level of description of social behavior and to gain insight into the relationship between scales, we asked whether the multitudinous processes associated with social behavior abide to general principles. In section "Classical Experimental Findings on Social Coordination" of this review, we presented a series of experimental snapshots taken from two levels (brain and behavior), all pointing to spatiotemporal metastability as a common organizing principle. Metastability arises from weak coupling (permissive of flexible binding) and diversity (tendency of the parts to unbind and act independently) – the tension between the two opens up a universe of complex coordinative patterns besides phase-locking. Metastability is especially well understood in its simpler forms near the border of the bifurcation from stable phase-locking (Kelso, 1991, 2012; Zhang et al., 2018). Furthermore, metastability probably remains in effect throughout more complex coordination patterns, even as we may fail to identify and quantify its more elusive forms (Tognoli and Kelso, 2014a; see also Zhang et al., 2020 for methodological advances using computational algebraic topology).

Recognizing that evolutionarily speaking, rhythm is a powerful way to put people in tune with one another (whether it be to dance, to compete, to make war, or to worship), we have taken oscillatory dynamics as the workhorse of our research paradigm. Yet, oscillatory behavior is all but its terminal end game. In section "A Chronology of the HKB Model's Augmentations," we laid out the state of our progress to bring the above experimental insights of social coordination into an evolving modeling framework: starting from a pair of symmetrical non-linearly coupled non-linear oscillators (**Figure 4**, eq. 1), crossing the crucial step of broken symmetry that unleashes metastability (eq. 2), and augmenting for discreteness (eq. 3), frequency adaptation (eq. 4), intentionally directed phase patterning (eq. 5), and finally scaling for multiple interacting agents (eq. 6) as a leap into a multiscale framework. These augmentations progressively evolve a repertoire of coordinative behaviors with increasing realism: elaborating on the intrinsic dynamics, coupling and within- and across-scale composition in a self-consistent manner. They echo numerous calls for models that reach beyond the original HKB model (Jirsa et al., 1998; Beek et al., 2002; Newell et al., 2008; Kelso et al., 2013; Avitabile et al., 2016; de Poel, 2016; Słowiński et al., 2018, 2020) and lay some foundation for generative approaches to the complexity of intrinsic and social dynamics that our interdisciplinary group continues to pursue.

Grounded in the dynamics of sensorimotor loops that couple perception and action between two or more individuals (**Figure 1**), more profound sociocognitive concepts quickly emerge. As posited in the introduction, an essential characteristic of the present modeling framework is that it approaches social behavioral dynamics from two standpoints: one for the intrinsic dynamics of the self and one for the coupling to the partners, their socialness. The equations governing the evolution of self-behavior (**Figure 4**, left hand-side of leftmost column) are dynamical mechanisms that intertwine self and others via the interaction (right hand-side of same). However, for the self to remain distinct from, yet informed by social partners, there needs to be a separation of time scales at which self-disposition and input from others influence individual behavior. Our model of social memory (**Figure 4**, eq. 4), contains distinct time scales and coupling for the moment-to-moment coordination of behavior (**Figure 4**, eq. 4, parameters a and 2b, purple color), and for the influence that the other(s) exert(s) more permanently on self-dispositions (**Figure 4**, eq. 4, parameter epsilon, red). If it is a key asset of biological adaptation to modify one's internal state (Maturana, 1970), then our work highlights how neurobehavioral symmetries at play in social interaction contribute to shaping human behavior (see also Dumas et al., 2012), a well-recognized concept in sociology, developmental psychology and learning science (Thelen and Smith, 1994; Mercer, 2011; Sheets-Johnstone, 2017).

From the intricacies of self and others (above) immediately follows the question of agency (Kelso, 2000, 2016; De Jaegher and Froese, 2009; Buhrmann and Di Paolo, 2017; Solfo et al., 2019). In a line of the model, we demonstrated how to stabilize initially unstable phase patterns using the equations for directed coordination (**Figure 4**, eq. 5; Dumas et al., 2014; Kostrubiec et al., 2015), which was transformed into an agentic learning tool in human machine interaction. On the subjective side, Kelso (2016) has theorized that a developmental phase transition to agency occurs when infants realize the impact of their action on the world. Kelso and Fuchs (2016) have developed a model of this phase transition. The question of agency reaches an apex of complexity when multiple intentionalities are conflated into a collective outcome. In our study of the Human Dynamic Clamp, we have shown that the model is able to tune various aspects of its coupling strength to modify the rate at which it converges to its "intention" and therefore gain or lose agency to a competitive human partner whose temporal dynamics is probably more constrained than the model's (Kelso et al., 2009; Dumas et al., 2014). A study of human brain (Dumas et al., 2020) suggests that the subjective sense of agency arising within such human-machine interactions has its root in the neural dynamics entrained by the movements from self and other, whose convergence occurs in the right parietal cortex. This set of results not only highlights the key role of right parietal areas in social coordination, but also points toward a link between sensorimotor neuromarkers and affective dimensions of human social cognition (see also Zhang et al., 2016), in agreement with paleocognitive accounts of the right hemisphere as an evolutionary neuroanatomical basis from predatory threat avoidance to social processing (Forrester and Todd, 2018). It is hoped that the present work may eventually speak to higher level processes such as the mentalizing versus simulation debate in social cognition (Gallagher, 2008; Frith and Frith, 2012; Sperduti et al., 2014; Alcalá-López et al., 2019).

A multiscale framework sufficiently mature to encompass sensory, cognitive and motor abilities will further allow one to explore the effect of traits and pathologies on coordinative competencies. We showed that the model cross-validates with experimental studies when parametric manipulations predictably induce phase transitions (i.e., a logic distinct from curve-fitting). The social coordination dynamics framework may overcome the curse (for the scientist, since functionally, it is a blessing) of behavioral "degeneracy" [equivalence of behavioral coordinative (dis)abilities arising under distinct individual sensorimotor organizations] by dissecting neurobehavioral roots of social behavior in conditions such as autism or Parkinson's Disease (see also Lagarde, 2013; Dodel et al., 2020 for related views). Specific experiments across traits and conditions, guided by modeling insights, also power neurobehavioral diagnostic tools with great specificity (e.g., Baillin et al., 2019).

Our most recent innovation with multiple agents complementing the prior dyadic formalism (**Figure 4**, eq. 6; Zhang et al., 2019) has provided a decisive stepping-stone for the multiscale framework that has long been envisioned. By marrying models of coordination based on statistical mechanics (the Kuramoto model, Kuramoto, 1984) and non-linear dynamics (extended HKB, Kelso et al., 1990), the Zhang et al. (2019) model from eq. 6 provides an experimentally validated framework where coordinative structures can exist within other coordinative structures – the ground zero for vertical integration across scales. From a complex systems perspective, this is a much-needed innovation because external control elements (the parameters that scientists tune and set) can now be incorporated by layering systems within systems, with the immediate consequence that the loose ends previously left in the hands of scientists can be returned to self-organizing principles and advance increasingly autonomous architectures recapitulating social behavior across scales. A neurocomputational model of social behavior (Dumas et al., 2012; Tognoli et al., 2018a) is but one of them. The development of this scalable, empirically validated framework also allows one to examine multiscale coordinative structures and study how they arise from simple (but no simpler) interaction between individuals. In particular, by introducing more "space" (degrees of freedom), this framework generalizes the impact of metastability, a mechanism originally discovered in dyadic interaction, to a system level: it creates spatiotemporal metastability, allowing a large-scale system to adopt very many different configurations in a sequential, recurrent manner. In other words, metastable coordination dynamics endows a system with an ability to generate complex, yet organized, spatiotemporal patterns – the sign of a true complex system.

AUTHOR CONTRIBUTIONS

All authors conceived the review presented in the manuscript. ET, MZ, CB, and JK contributed to the final manuscript.

FUNDING

The authors of this work are supported by NIMH Grant MH080838, NIBIB Grant EB025819, the FAU Foundation, FAU I-SENSE, and FAU Brain Institute.

REFERENCES

Abbott, A. E., Linke, A. C., Nair, A., Jahedi, A., Alba, L. A., Keown, C. L., et al. (2017). Repetitive behaviors in autism are linked to imbalance of corticostriatal connectivity: a functional connectivity MRI study. *Soc. Cogn. Affect. Neurosci.* 13, 32–42. doi: 10.1093/scan/nsx129

Adolphs, R. (2009). The social brain: neural basis of social knowledge. *Annu. Rev. Psychol.* 60, 693–716. doi: 10.1146/annurev.psych.60.110707.163514

Alcalá-López, D., Vogeley, K., Binkofski, F., and Bzdok, D. (2019). Building blocks of social cognition: mirror, mentalize, share? *Cortex* 118, 4–18. doi: 10.1016/j.cortex.2018.05.006

Anders, S., Heinzle, J., Weiskopf, N., Ethofer, T., and Haynes, J. D. (2011). Flow of affective information between communicating brains. *Neuroimage* 54, 439–446. doi: 10.1016/j.neuroimage.2010.07.004

Anderson, P. W. (1972). More is different. *Science* 177, 393–396.

Arató, M. (2003). A famous nonlinear stochastic equation (Lotka-Volterra model with diffusion). *Math. Comput. Model.* 38, 709–726. doi: 10.1016/s0895-7177(03)90056-2

Attwell, D., and Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow Metabol.* 21, 1133–1145. doi: 10.1097/00004647-200110000-00001

Avitabile, D., Słowiński, P., Bardy, B., and Tsaneva-Atanasova, K. (2016). Beyond in-phase and anti-phase coordination in a model of joint action. *Biol. Cybern.* 110, 201–216. doi: 10.1007/s00422-016-0691-9

Axelrod, R. (1997). *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*. Princeton, NJ: Princeton University Press.

Babiloni, F., and Astolfi, L. (2014). Social neuroscience and hyperscanning techniques: past, present and future. *Neurosci. Biobehav. Rev.* 44, 76–93. doi: 10.1016/j.neubiorev.2012.07.006

Baillin, F., Lefebvre, A., Pedoux, A., Beauxis, Y., Engemann, D., Maruani, A., et al. (2019). Interactive psychometrics for autism with the human dynamic clamp: interpersonal synchrony from sensory-motor to socio-cognitive domains. *medRxiv [Perprint]* doi: 10.1101/19013771

Balconi, M., and Vanutelli, M. E. (2017). Cooperation and competition with hyperscanning methods: review and future application to emotion domain. *Front. Comput. Neurosci.* 11:86. doi: 10.3389/fncom.2017.00086

Banerjee, A., Tognoli, E., Kelso, J. A. S., and Jirsa, V. K. (2012). Spatiotemporal reorganization of large-scale neural assemblies mediates bimanual coordination. *NeuroImage* 62, 1582–1592. doi: 10.1016/j.neuroimage.2012.05.046

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.* 59, 617–645.

Batty, M. (2007). *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals*. Cambridge, MA: The MIT press.

Beek, P. J., Peper, C. E., and Daffertshofer, A. (2002). Modeling rhythmic interlimb coordination: beyond the Haken–Kelso–Bunz model. *Brain Cogn.* 48, 149–165. doi: 10.1006/brcg.2001.1310

Bennett, M., Schatz, M. F., Rockwood, H., and Wiesenfeld, K. (2002). Huygens's clocks. *Proc. R. Soc. London. Ser. A.* 458, 563–579.

Bonabeau, E. (2002). Agent-based modeling: methods and techniques for simulating human systems. *Proc. Natl. Acad. Sci. U.S.A.* 99(Suppl. 3), 7280–7287. doi: 10.1073/pnas.082080899

Bressler, S. L., and Kelso, J. A. S. (2001). Cortical coordination dynamics and cognition. *Trends Cogn. Sci.* 5, 26–36. doi: 10.1016/s1364-6613(00)01564-3

Bressler, S. L., and Tognoli, E. (2006). Operational principles of neurocognitive networks. *Int. J. Psychophysiol.* 60, 139–148. doi: 10.1016/j.ijpsycho.2005.12.008

Brown, P. (2003). Oscillatory nature of human basal ganglia activity: relationship to the pathophysiology of Parkinson's disease. *Mov. Disord.* 18, 357–363. doi: 10.1002/mds.10358

Buhrmann, T., and Di Paolo, E. (2017). The sense of agency–a phenomenological consequence of enacting sensorimotor schemes. *Phenomenol. Cogn. Sci.* 16, 207–236. doi: 10.1007/s11097-015-9446-7

Buzsáki, G. (2010). Neural syntax: cell assemblies, synapsembles, and readers. *Neuron* 68, 362–385. doi: 10.1016/j.neuron.2010.09.023

Cacioppo, S., Zhou, H., Monteleone, G., Majka, E. A., Quinn, K. A., Ball, A. B., et al. (2014). You are in sync with me: neural correlates of interpersonal synchrony with a partner. *Neuroscience* 277, 842–858. doi: 10.1016/j.neuroscience.2014.07.051

Castellano, C., Fortunato, S., and Loreto, V. (2009). Statistical physics of social dynamics. *Rev. Mod. Phys.* 81:591. doi: 10.1103/revmodphys.81.591

Coey, C. A., Varlet, M., and Richardson, M. J. (2012). Coordination dynamics in a socially situated nervous system. *Front. Hum. Neurosci.* 6:164. doi: 10.3389/fnhum.2012.00164

D'Ausilio, A., Novembre, G., Fadiga, L., and Keller, P. E. (2015). What can music tell us about social interaction? *Trends Cogn. Sci.* 19, 111–114. doi: 10.1016/j.tics.2015.01.005

De Jaegher, H., and Froese, T. (2009). On the role of social interaction in individual agency. *Adapt. Behav.* 17, 444–460. doi: 10.1177/1059712309343822

De Luca, C., Jantzen, K. J., Comani, S., Bertollo, M., and Kelso, J. A. S. (2010). Striatal activity during intentional switching depends on pattern stability. *J. Neurosci.* 30, 3167–3174. doi: 10.1523/jneurosci.2673-09.2010

de Poel, H. J. (2016). Anisotropy and antagonism in the coupling of two oscillators: concepts and applications for between-person coordination. *Front. Psychol.* 7:1947. doi: 10.3389/fpsyg.2016.01947

Dodel, S., Tognoli, E., and Kelso, J. A. S. (2020). Degeneracy and complexity in neuro-behavioral correlates of team coordination. *Front. Hum. Neurosci. Cog. Neurosci.* doi: 10.3389/fnhum.2020.00328

Duarte, R., Araújo, D., Davids, K., Travassos, B., Gazimba, V., and Sampaio, J. (2012). Interpersonal coordination tendencies shape 1-vs-1 sub-phase performance outcomes in youth soccer. *J. Sports Sci.* 30, 871–877. doi: 10.1080/02640414.2012.675081

Dumas, G., Chavez, M., Nadel, J., and Martinerie, J. (2012). Anatomical connectivity influences both intra-and inter-brain synchronizations. *PLoS One* 7:e36414. doi: 10.1371/journal.pone.0036414

Dumas, G., de Guzman, G. C., Tognoli, E., and Kelso, J. A. S. (2014). The human dynamic clamp as a paradigm for social interaction. *Proc. Natl. Acad. Sci. U.S.A.* 111, E3726–E3734.

Dumas, G., Lachat, F., Martinerie, J., Nadel, J., and George, N. (2011). From social behaviour to brain synchronization: review and perspectives in hyperscanning. *IRBM* 32, 48–53. doi: 10.1016/j.irbm.2011.01.002

Dumas, G., Lefebvre, A., Zhang, M., Tognoli, E., and Kelso, J. A. S. (2018). The human dynamic clamp: a probe for coordination across neural, behavioral, and social scales. In *Complexity and Synergetics*. eds Müller S., Plath P., Radons G., Fuchs A. (Berlin: Springer), 317–332. doi: 10.1007/978-3-319-64334-2_24

Dumas, G., Moreau, Q., Tognoli, E., and Kelso, J. A. S. (2020). The human dynamic clamp reveals the fronto-parietal network linking real-time social coordination and cognition. *Cereb. Cortex* 30, 3271–3285. doi: 10.1093/cercor/bhz308

Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., and Garnero, L. (2010). Inter-brain synchronization during social interaction. *PLoS One* 5:e12166. doi: 10.1371/journal.pone.0012166

Dunbar, R. I., and Shultz, S. (2007). Evolution in the social brain. *Science* 317, 1344–1347.

Edelman, G. E., and Mountcastle, V. B. (1978). *The Mindful Brain*. Cambridge, MA: MIT Press.

Fine, J. M., and Amazeen, E. L. (2014). Stabilizing perceptual-motor asymmetries during social coordination. *Hum. Mov. Sci.* 34, 91–108. doi: 10.1016/j.humov.2014.01.004

Fine, J. M., Likens, A. D., Amazeen, E. L., and Amazeen, P. G. (2015). Emergent complexity matching in interpersonal coordination: local dynamics and global variability. *J. Exp. Psychol.* 41, 723–737. doi: 10.1037/xhp0000046

Forrester, G. S., and Todd, B. K. (2018). A comparative perspective on lateral biases and social behavior. *Prog. Brain Res.* 238, 377–403. doi: 10.1016/bs.pbr.2018.06.014

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., and Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9673–9678. doi: 10.1073/pnas.0504136102

Freeman, W. J. (1980). Use of spatial deconvolution to compensate for distortion of EEG by volume conduction. *IEEE Trans. Biomed. Eng.* 27, 421–429. doi: 10.1109/tbme.1980.326750

Friston, K. (2011). What is optimal about motor control? *Neuron* 72, 488–498. doi: 10.1016/j.neuron.2011.10.018

Frith, C. D., and Frith, U. (2012). Mechanisms of social cognition. *Annu. Rev. Psychol.* 63, 287–313.

Fuchs, A. (2014). *Nonlinear Dynamics in Complex Systems. Theory and Applications for the Life-, Neuro- and Natural Sciences*. Berlin: Springer.

Fuchs, A., Jirsa, V. K., Haken, H., and Kelso, J. A. S. (1996). Extending the HKB model of coordinated movement to oscillators with different eigenfrequencies. *Biol. Cybern.* 74, 21–30. doi: 10.1007/s004220050215

Fuchs, A., Jirsa, V. K., and Kelso, J. A. S. (2000a). Theory of the relation between human brain activity (MEG) and hand movements. *Neuroimage* 11, 359–369. doi: 10.1006/nimg.1999.0532

Fuchs, A., Mayville, J., Cheyne, D., Weinberg, H., Deecke, L., and Kelso, J. A. S. (2000b). Spatiotemporal analysis of neuromagnetic events underlying the emergence of coordinative instabilities. *NeuroImage* 12, 71–84. doi: 10.1006/nimg.2000.0589

Fuchs, A., and Kelso, J. A. S. (1994). A theoretical note on models of interlimb coordination. *J. Exp. Psychol.* 20, 1088–1097. doi: 10.1037/0096-1523.20.5.1088

Fuchs, A., and Kelso, J. A. S. (2009). "Movement coordination," in *Encyclopedia of Complexity and Systems Science*, ed. R. A. Meyers (New York, NY: Springer) 5718–5736.

Fuchs, A., and Kelso, J. A. S. (2018). "Coordination dynamics and synergetics: from finger movements to brain patterns and ballet dancing," in *Complexity and Synergetics* (New York, NY: Springer), 301–316. doi: 10.1007/978-3-319-64334-2_23

Fuchs, A., Kelso, J. A. S., and Haken, H. (1992). Phase transitions in the human brain: spatial mode dynamics. *Int. J. Bifurcat. Chaos* 2, 917-939.

Fusaroli, R., Rączaszek-Leonardi, J., and Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas Psychol.* 32, 147–157. doi: 10.1016/j.newideapsych.2013.03.005

Gal, A., Eytan, D., Wallach, A., Sandler, M., Schiller, J., and Marom, S. (2010). Dynamics of excitability over extended timescales in cultured cortical neurons. *J. Neurosci.* 30, 16332–16342. doi: 10.1523/jneurosci.4859-10.2010

Gallagher, S. (2008). Inference or interaction: social cognition without precursors. *Phil. Explor.* 11, 163–174. doi: 10.1080/13869790802239227

Gell-Mann, M. (1988). Simplicity and complexity in the description of nature. *Eng. Sci.* 51, 2–9.

Gilbert, N., and Terna, P. (2000). How to build and use agent-based models in social science. *Mind Soc.* 1, 57–72. doi: 10.1007/bf02512229

Gintis, H. (2013). Markov models of social dynamics: theory and applications. *ACM TIST* 4, 53–51.

Glass, L. (2001). Synchronization and rhythmic processes in physiology. *Nature* 410:277. doi: 10.1038/35065745

Glass, L., and MacKey, M. C. (1988). *From Clocks to Chaos: The rhythms of life*. Princeton, NJ: Princeton University Press.

Gray, C. M. (1994). Synchronous oscillations in neuronal systems: mechanisms and functions. *J. Comput. Neurosci.* 1, 11–38. doi: 10.1007/bf00962716

Grillner, S., Ekeberg, Ö, El Manira, A., Lansner, A., Parker, D., Tegner, J., et al. (1998). Intrinsic function of a neuronal network—a vertebrate central pattern generator. *Brain Res. Rev.* 26, 184–197. doi: 10.1016/s0165-0173(98)00002-2

Haken, H. (1977). *Synergetics, An Introduction: Non-Equilibrium Phase Transitions and Self-Organization in Physics, Chemistry and Biology*. Berlin: Springer.

Haken, H. (2013). *Principles of Brain Functioning: A Synergetic Approach to Brain Activity, Behavior and Cognition*. Berlin: Springer Science & Business Media.

Haken, H., Kelso, J. A. S., and Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biol. Cybern.* 51, 347–356. doi: 10.1007/bf00336922

Hari, R., and Kujala, M. V. (2009). Brain basis of human social interaction: from concepts to brain imaging. *Physiol. Rev.* 89, 453–479. doi: 10.1152/physrev.00041.2007

Harris, A. Z., and Gordon, J. A. (2015). Long-range neural synchrony in behavior. *Annu. Rev. Neurosci.* 38, 171–194. doi: 10.1146/annurev-neuro-071714-034111

Hegselmann, R. (1998). "Modeling social dynamics by cellular automata," in *Computer Modeling of Network Dynamics*, eds W. B. G. Liebrand, A. Nowak, and R. Hegselman (London: Sage Publication), 37–64.

Hirata, M., Ikeda, T., Kikuchi, M., Kimura, T., Hiraishi, H., Yoshimura, Y., et al. (2014). Hyperscanning MEG for understanding mother–child cerebral interactions. *Front. Hum. Neurosci.* 8:118. doi: 10.3389/fnhum.2014.00118

Irsoy, O., and Cardie, C. (2014). "Opinion mining with deep recurrent neural networks," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (Hong Kong: EMNLP), 720–728.

Issartel, J., Gueugnon, M., and Marin, L. (2017). Understanding the impact of expertise in joint and solo-improvisation. *Front. Psychol.* 8:1078. doi: 10.3389/fpsyg.2017.01078

Janata, P., Tomic, S. T., and Haberman, J. M. (2012). Sensorimotor coupling in music and the psychology of the groove. *J. Exp. Psychol.* 141:54. doi: 10.1037/a0024208

Jantzen, K. J., and Kelso, J. A. S. (2007). "Neural coordination dynamics of human sensorimotor behavior: a review," in *Handbook of Brain Connectivity*, eds V. K. Jirsa, A. McIntosh (Berlin: Springer), 421–461. doi: 10.1007/978-3-540-71512-2_15

Jantzen, K. J., Oullier, O., and Kelso, J. A. S. (2008). Neuroimaging coordination dynamics in the sport sciences. *Methods* 45, 325–335. doi: 10.1016/j.ymeth.2008.06.001

Jantzen, K. J., Steinberg, F. L., and Kelso, J. A. S. (2004). Brain networks underlying human timing behavior are influenced by prior context. *Proc. Natl. Acad. Sci. U.S.A.* 101, 6815–6820. doi: 10.1073/pnas.0401300101

Jantzen, K. J., Steinberg, F. L., and Kelso, J. A. S. (2009). Coordination dynamics of large-scale neural circuitry underlying rhythmic sensorimotor behavior. *J. Cogn. Neurosci.* 21, 2420–2433. doi: 10.1162/jocn.2008.21182

Jeka, J. J., Kelso, J. A. S., and Kiemel, T. (1993). Pattern switching in human multilimb coordination dynamics. *Bull. Math. Biol.* 55, 829–845. doi: 10.1007/bf02460675

Jirsa, V. K., Fuchs, A., and Kelso, J. A. S. (1998). Connecting cortical and behavioral dynamics: bimanual coordination. *Neural. Comput.* 10, 2019–2045. doi: 10.1162/089976698300016954

Jirsa, V. K., and Haken, H. (1997). A derivation of a macroscopic field theory of the brain from the quasi-microscopic neural dynamics. *Phys. D* 99, 503–526. doi: 10.1016/s0167-2789(96)00166-2

Jirsa, V. K., and Kelso, J. A. S. (2005). The excitator as a minimal model for the coordination dynamics of discrete and rhythmic movement generation. *J. Mot. Behav.* 37, 35–51. doi: 10.3200/jmbr.37.1.35-51

Kawasaki, M., Kitajo, K., and Yamaguchi, Y. (2018). Sensory-motor synchronization in the brain corresponds to behavioral synchronization between individuals. *Neuropsychologia* 119, 59–67. doi: 10.1016/j.neuropsychologia.2018.07.026

Kawasaki, M., Yamada, Y., Ushiku, Y., Miyauchi, E., and Yamaguchi, Y. (2013). Inter-brain synchronization during coordination of speech rhythm in human-to-human social interaction. *Sci. Rep.* 3:1692.

Keller, P. E., Novembre, G., and Hove, M. J. (2014). Rhythm in joint action: psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Phil. Trans. R. Soc. B Biol. Sci.* 369:20130394. doi: 10.1098/rstb.2013.0394

Kelso, J. A. S. (1981). On the oscillatory basis of movement. *Bull. Psychon. Soc.* 18: 63.

Kelso, J. A. S. (1984). Phase transitions and critical behavior in human bimanual coordination. *Am. J. Physiol.* 246, R1000–R1004.

Kelso, J. A. S. (1991). "Behavioral and neural pattern generation: the concept of neurobehavioral dynamical system (n.d.)," in *Cardiorespiratory and Motor Coordination*, eds H. P. Koepchen and T. Huopaniemi (Berlin: Springer), 224–238. doi: 10.1007/978-3-642-75507-1_25

Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: MIT press.

Kelso, J. A. S. (2000). "Principles of dynamic pattern formation and change for a science of human behavior," in *Developmental Science and the Holistic Approach*, eds L. R. Bergman, R. B. Cairns, L.-G. Nilsson, and L. Nystedt (Abingdon: Routledge), 73–94.

Kelso, J. A. S. (2001). Metastable coordination dynamics of brain and behavior. *Brain Neural Netw.* 8, 125–130. doi: 10.3902/jnns.8.125

Kelso, J. A. S. (2008). Haken-Kelso-Bunz model. *Scholarpedia* 3:1612. doi: 10.4249/scholarpedia.1612

Kelso, J. A. S. (2009a). "Coordination Dynamics," in *Encyclopedia of Complexity and System Science*, ed. R. A. Meyers (Berlin: Springer), 1537–1564.

Kelso, J. A. S. (2009b). Synergies: atoms of brain and behavior. *Adv. Exp. Med. Biol.* 629, 83–91. doi: 10.1007/978-0-387-77064-2_5

Kelso, J. A. S. (2010). Instabilities and phase transitions in human brain and behavior. *Front. Hum. Neurosci.* 4:23. doi: 10.3389/fnhum.2010.00023

Kelso, J. A. S. (2012). Multistability and metastability: understanding dynamic coordination in the brain. *Phil. Trans. R. Soci. B.* 367, 906–918. doi: 10.1098/rstb.2011.0351

Kelso, J. A. S. (2016). On the self-organizing origins of agency. *Trends Cogn. Sci.* 20, 490–499. doi: 10.1016/j.tics.2016.04.004

Kelso, J. A. S., Bressler, S. L., Buchanan, S., DeGuzman, G. C., Ding, M., Fuchs, A., et al. (1991). "Cooperative and critical phenomena in the human brain revealed by multiple squids," in *Measuring Chaos in the Human Brain*, eds D. Duke and W. Pritchard (New Jersey, NJ: World Scientific), 97-112.

Kelso, J. A. S., Bressler, S. L., Buchanan, S., DeGuzman, G. C., Ding, M., Fuchs, A., et al. (1992). A phase transition in human brain and behavior. *Phys. Lett. A* 169, 134–144. doi: 10.1016/0375-9601(92)90583-8

Kelso, J. A. S., de Guzman, G. C., Reveley, C., and Tognoli, E. (2009). Virtual partner interaction (VPI): exploring novel behaviors via coordination dynamics. *PLoS One* 4:e5749. doi: 10.1371/journal.pone.0005749

Kelso, J. A. S., Del Colle, J. D., and Schöner, G. (1990). *Action-Perception as a Pattern Formation Process. Attention and Performance XIII*. Hillsdale, NJ: Erlbaum, 139-169.

Kelso, J. A. S., Dumas, G., and Tognoli, E. (2013). Outline of a general theory of behavior and brain coordination. *Neural Netw*. 37, 120–131. doi: 10.1016/j.neunet.2012.09.003

Kelso, J. A. S., and Fuchs, A. (2016). The coordination dynamics of mobile conjugate reinforcement. *Biol. Cybern.* 110, 41–53. doi: 10.1007/s00422-015-0676-0

Kelso, J. A. S., Fuchs, A., and Jirsa, V. K. (1999). "Traversing scales of brain and behavioral organization I: concepts and experiments," in *Analysis of Neurophysiological Brain Functioning*, eds C. Uhl (Berlin: Springer), 73–89. doi: 10.1007/978-3-642-60007-4_4

Kelso, J. A. S., Fuchs, A., Lancaster, R., Holroyd, T., Cheyne, D., and Weinberg, H. (1998). Dynamic cortical activity in the human brain reveals motor equivalence. *Nature* 392, 814–818. doi: 10.1038/33922

Kelso, J. A. S., and Haken, H. (1995). "New laws to be expected in the organism: synergetics of brain and behavior," in *What is Life? The Next 50 Years*, eds M. Murphy and L. O'Neill (Cambridge: Cambridge University Press), 137–160. doi: 10.1017/cbo9780511623295.012

Kelso, J. A. S., Holt, K. G., Rubin, P., and Kugler, P. N. (1981). Patterns of human interlimb coordination emerge from the properties of non-linear oscillatory processes: Theory and data. *J. Mot. Behav.* 13, 226–261. doi: 10.1080/00222895.1981.10735251

Kelso, J. A. S., Schöner, G., Scholz, J. P., and Haken, H. (1987). Phase-locked modes, phase transitions and component oscillators in coordinated biological motion, *Phys. Scr.* 35, 79–87. doi: 10.1088/0031-8949/35/1/020

Kelso, J. A. S., Southard, D., and Goodman, D. (1979). On the nature of human interlimb coordination. *Science* 203, 1029–1031. doi: 10.1126/science.424729

Kelso, J. A. S., and Tognoli, E. (2007). "Toward a complementary neuroscience: metastable coordination dynamics of the brain," in *Neurodynamics of Cognition and Consciousness*, eds R. Kozma and L. Perlovsky (Berlin: Springer).

Kelso, J. A. S., and Tuller, B. (1984). "A dynamical basis for action systems," in *Handbook of Cognitive Neuroscience*, ed. M. S. Gazzaniga (New York, NY: Plenum), 321-356.

Kelso, J. A. S., and Zanone, P. G. (2002). Coordination dynamics of learning and transfer across different effector systems. *J. Exp. Psychol.* 28, 776–797. doi: 10.1037/0096-1523.28.4.776

Khazipov, R., and Luhmann, H. J. (2006). Early patterns of electrical activity in the developing cerebral cortex of humans and rodents. *Trends Neurosci.* 29, 414–418. doi: 10.1016/j.tins.2006.05.007

Kimura, M., Saito, K., Nakano, R., and Motoda, H. (2010). Extracting influential nodes on a social network for information diffusion. *Data Min. Knowl. Disc.* 20:70. doi: 10.1007/s10618-009-0150-5

Koike, T., Tanabe, H. C., and Sadato, N. (2015). Hyperscanning neuroimaging technique to reveal the "two-in-one" system in social interactions. *Neurosci. Res.* 90, 25–32. doi: 10.1016/j.neures.2014.11.006

Konvalinka, I., and Roepstorff, A. (2012). The two-brain approach: how can mutually interacting brains teach us something about social interaction? *Front. Hum. Neurosci.* 6:215. doi: 10.3389/fnhum.2012.00215

Kostrubiec, V., Dumas, G., Zanone, P. G., and Kelso, J. A. S. (2015). The virtual teacher (VT) paradigm: learning new patterns of interpersonal coordination using the human dynamic clamp. *PLoS One* 10:e0142029. doi: 10.1371/journal.pone.0142029

Kostrubiec, V., Zanone, P.-G., Fuchs, A., and Kelso, J. A. S. (2012). Beyond the blank slate: routes to learning new coordination patterns depend on the intrinsic dynamics of the learner —experimental evidence and theoretical model. *Front. Hum. Neurosci.* 6:212. doi: 10.3389/fnhum.2012.00222

Kramer, D. L., and McLaughlin, R. L. (2001). The behavioral ecology of intermittent locomotion. *Am. Zool.* 41, 137–153. doi: 10.1093/icb/41.2.137

Kuramoto, Y. (1984). *Chemical Oscillations, Waves, and Turbulence*. Berlin: Springer.

Lagarde, J. (2013). Challenges for the understanding of the dynamics of social coordination. *Front. Neurorobot.* 7:18. doi: 10.3389/fnbot.2013.00018

Lagarde, J., DeGuzman, G. C., Oullier, O., and Kelso, J. A. S. (2006). Interpersonal interactions during boxing: data and model. *J. Sport Exerc. Psychol.* 28, S108–S108.

Laughlin, R. B., and Pines, D. (2000). The theory of everything. *Proc. Nat. Acad. Sci. U.S.A.* 97, 28–31.

Liu, P., and Ma, T. H. (2018). "Strong structural controllability of multi-agent systems with switching topologies," in *Proceedings of the 2018 10th International Conference on Modelling, Identification and Control (ICMIC)* (Piscataway, NJ: IEEE), 1–6.

López-Pintado, D. (2008). Diffusion in complex social networks. *Games Econ. Behav*. 62, 573–590.

MacMahon, J. A., Phillips, D. L., Robinson, J. V., and Schimpf, D. J. (1978). Levels of biological organization: an organism-centered approach. *BioScience* 28, 700–704. doi: 10.2307/1307320

Macrae, C. N., and Miles, L. K. (2012). "Revisiting the sovereignty of social cognition: finally some action," in *The SAGE handbook of Social Cognition*, eds S. T. Fiske and C. N. Macrae (Thousand Oaks, CA: Sage), 1–11. doi: 10.4135/9781446247631.n1

Marom, S. (2010). Neural timescales or lack thereof. *Prog. Neurobiol.* 90, 16–28. doi: 10.1016/j.pneurobio.2009.10.003

Marsh, K. L., Richardson, M. J., and Schmidt, R. C. (2009). Social connection through joint action and interpersonal coordination. *Top. Cogn. Sci.* 1, 320–339. doi: 10.1111/j.1756-8765.2009.01022.x

Matsuda, H., Ogita, N., Sasaki, A., and Satô, K. (1992). Statistical mechanics of population: the lattice Lotka-Volterra model. *Prog. Theor. Phys.* 88, 1035–1049. doi: 10.1143/ptp/88.6.1035

Maturana, H. R. (1970). *Biology of cognition*. Urbana: University of Illinois, 4–9.

Mayville, J. M., Bressler, S. L., Fuchs, A., and Kelso, J. A. S. (1999). Spatiotemporal reorganization of electrical activity in the human brain associated with a timing transition. *Exp. Brain Res.* 127, 371–381. doi: 10.1007/s002210050805

Mayville, J. M., Jantzen, K. J., Fuchs, A., Steinberg, F. L., and Kelso, J. A. S. (2002). Cortical and subcortical networks underlying syncopated and synchronized coordination revealed using fMRI. *Hum. Brain Mapp.* 17, 214–229. doi: 10.1002/hbm.10065

Mercer, S. (2011). The self as a complex dynamic system. *Stud. Sec. Lang. Learn. Teach.* 1, 57–82.

Meyer-Lindenberg, A., Zieman, U., Hajak, G., Cohen, L., and Berman, K. (2002). Transition between dynamical states of differing stability in the human brain. *Proc. Natl. Acad. Sci. U.S.A.* 99, 10948–10953. doi: 10.1073/pnas.162114799

Minorsky, N. (1962). *Nonlinear oscillations*. Princeton, NJ: Van Nostrand.

Moreau, Q., Galvan, L., Nazir, T. A., and Paulignan, Y. (2016). Dynamics of social interaction: kinematic analysis of a joint action. *Front. Psychol.*7:2016. doi: 10.3389/fpsyg.2016.02016

Morowitz, H. J. (1968). *Energy Flow in Biology; Biological Organization as a Problem in Thermal Physics*. New York, NY: Academic Press.

Mu, Y., Cerritos, C., and Khan, F. (2018). Neural mechanisms underlying interpersonal coordination: a review of hyperscanning research. *Soc. Personal. Psychol. Compass* 12:e12421. doi: 10.1111/spc3.12421

Naeem, M., Prasad, G., Watson, D. R., and Kelso, J. A. S. (2012). Electrophysiological signatures of intentional social coordination in the 10–12 Hz range. *Neuroimage* 59, 1795–1803. doi: 10.1016/j.neuroimage.2011.08.010

Nalepka, P., Kallen, R. W., Chemero, A., Saltzman, E., and Richardson, M. J. (2017). Herd those sheep: emergent multiagent coordination and behavioral mode switching. *Psychol. Sci.* 28, 630–650. doi: 10.1177/0956797617692107

Nalepka, P., Lamb, M., Kallen, R. W., Shockley, K., Chemero, A., Saltzman, E., et al. (2019). Human social motor solutions for human-machine interaction in dynamical task contexts. *Proc. Natl. Acad. Sci. U.S.A.* 116, 1437–1446. doi: 10.1073/pnas.1813164116

Newell, K. M., Liu, Y. T., and Mayer-Kress, G. (2008). Landscapes beyond the HKB model. In *Coordination: Neural, behavioral and Social Dynamics*, eds A. Fuchs, V.K. Jirsa (Berlin: Springer), 27–44. doi: 10.1007/978-3-540-74479-5_2

Nolte, G., Bai, O., Wheaton, L., Mari, Z., Vorbach, S., and Hallett, M. (2004). Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clin. Neurophysiol.* 115, 2292–2307. doi: 10.1016/j.clinph.2004.04.029

Nordham, C. A., Tognoli, E., Fuchs, A., and Kelso, J. A. S. (2018). How interpersonal coordination affects individual behavior (and vice versa):

experimental analysis and adaptive HKB model of social memory. *Ecol. Psychol.* 30, 224–249. doi: 10.1080/10407413.2018.1438196

Nowak, A. (2004). Dynamical minimalism: why less is more in psychology. *Pers. Soc. Psychol. Rev.* 8, 183–192. doi: 10.1207/s15327957pspr0802_12

Nunez, P. L., Srinivasan, R., Westdorp, A. F., Wijesinghe, R. S., Tucker, D. M., Silberstein, R. B., et al. (1997). EEG coherency: I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. *Electroencephalogr. Clin. Neurophysiol.* 103, 499–515.

Oullier, O., De Guzman, G. C., Jantzen, K. J., Lagarde, J., and Kelso, J. A. S. (2008). Social coordination dynamics: measuring human bonding. *Soc. Neurosci.* 3, 178–192. doi: 10.1080/17470910701563392

Oullier, O., Jantzen, K. J., Steinberg, F. L., and Kelso, J. A. S. (2004). Neural substrates of real and imagined sensorimotor coordination. *Cereb. Cortex* 15, 975–985. doi: 10.1093/cercor/bhh198

Oullier, O., and Kelso, J. A. S. (2009). "Social coordination from the perspective of coordination dynamics," in *Encyclopedia of Complexity and Systems Science*, ed. R. A. Meyers (Berlin: Springer), 8198–8212.

Pascual-Marqui, R. D. (2007). Coherence and phase synchronization: generalization to pairs of multivariate time series, and removal of zero-lag contributions. *arXiv* [Preprint]. https://arxiv.org/abs/0706.1776 (accessed October 12, 2019).

Pérez, A., Carreiras, M., and Duñabeitia, J. A. (2017). Brain-to-brain entrainment: EEG interbrain synchronization while speaking and listening. *Sci. Rep.* 7:4190.

Pezzulo, G., Donnarumma, F., Dindo, H., D'Ausilio, A., Konvalinka, I., and Castelfranchi, C. (2019). The body talks: sensorimotor communication and its brain and kinematic signatures. *Phys. Life Rev.* 28, 1–21. doi: 10.1016/j.plrev.2018.06.014

Pfeiffer, U. J., Vogeley, K., and Schilbach, L. (2013). From gaze cueing to dual eye-tracking: novel approaches to investigate the neural correlates of gaze in social interaction. *Neurosci. Biobehav. Rev.* 37, 2516–2528. doi: 10.1016/j.neubiorev.2013.07.017

Pikovsky, A., Rosenblum, M., and Kurths, J. (2002). *Synchronization: A Universal Concept in Nonlinear Science*. Cambridge: Cambridge University Press.

Reddish, P., Fischer, R., and Bulbulia, J. (2013). Let's dance together: synchrony, shared intentionality and cooperation. *PLoS One* 8:e71182. doi: 10.1371/journal.pone.0071182

Reindl, V., Gerloff, C., Scharke, W., and Konrad, K. (2018). Brain-to-brain synchrony in parent-child dyads and the relationship with emotion regulation revealed by fNIRS-based hyperscanning. *Neuroimage* 178, 493–502. doi: 10.1016/j.neuroimage.2018.05.060

Richardson, M. J., Dale, R., and Marsh, K. L. (2014). "Complex dynamical systems in social and personality psychology," in *Handbook of Research Methods in Social and Personality Psychology*, eds H. T. Reis, and C. M. Judd (New York, NY: Cambridge University Press) 253. doi: 10.1017/cbo9780511996481.015

Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R., and Schmidt, R. C. (2007). Rocking together: dynamics of intentional and unintentional interpersonal coordination. *Hum. Mov. Sci.* 26, 867–891. doi: 10.1016/j.humov.2007.07.002

Righetti, L., Buchli, J., and Ijspeert, A. J. (2005). "From dynamic hebbian learning for oscillators to adaptive central pattern generators," in *Proceedings of 3rd International Symposium on Adaptive Motion in Animals and Machines–AMAM 2005 (No. CONF)* (Ilmenau: Verlag ISLE).

Righetti, L., Buchli, J., and Ijspeert, A. J. (2006). Dynamic hebbian learning in adaptive frequency oscillators. *Phys. D* 216, 269–281. doi: 10.1016/j.physd.2006.02.009

Robertson, S. S. (1993). Oscillation and complexity in early infant behavior. *Child Dev.* 64, 1022–1035.

Sänger, J., Lindenberger, U., and Müller, V. (2011). Interactive brains, social minds. *Commun. Integr. Biol.* 4, 655–663. doi: 10.4161/cib.17934

Schmidt, R. C., Carello, C., and Turvey, M. T. (1990). Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people. *J. Exp. Psychol.* 16, 227–247. doi: 10.1037/0096-1523.16.2.227

Schmidt, R. C., Fitzpatrick, P., Caron, R., and Mergeche, J. (2011). Understanding social motor coordination. *Hum. Mov. Sci.* 30, 834–845. doi: 10.1016/j.humov.2010.05.014

Schmidt, R. C., and Richardson, M. J. (2008). "Dynamics of interpersonal coordination," in *Coordination: Neural, Behavioral and Social Dynamics*, eds A. Fuchs and V. K. Jirsa (Berlin: Springer), 281–308. doi: 10.1007/978-3-540-74479-5_14

Schöner, G., Jiang, W. Y., and Kelso, J. A. S. (1990). A synergetic theory of quadrupedal gaits and gait transitions. *J. Theor. Biol.* 142, 359–391. doi: 10.1016/s0022-5193(05)80558-2

Schöner, G., and Kelso, J. A. S. (1988a). A synergetic theory of environmentally-specified and learned patterns of movement coordination. *Biol. Cybern.* 58, 71–80. doi: 10.1007/bf00364153

Schöner, G., and Kelso, J. A. S. (1988b). Dynamic pattern generation in behavioral and neural systems. *Science* 239, 1513–1520. doi: 10.1126/science.3281253

Schweitzer, F. (2007). *Brownian Agents and Active Particles: Collective Dynamics in the Natural and Social Sciences*. Berlin: Springer.

Sheets-Johnstone, M. (2017). Agency: phenomenological insights and dynamic complementarities. *Humanist. Psychol.* 45, 1–22. doi: 10.1037/hum0000058

Singer, B., and Spilerman, S. (1976). The representation of social processes by Markov models. *Am. J. Sociol.* 82, 1–54. doi: 10.1086/226269

Singer, W. (1999). Neuronal synchrony: a versatile code for the definition of relations? *Neuron* 24, 49–65. doi: 10.1016/s0896-6273(00)80821-1

Słowiński, P., Al-Ramadhani, S., and Tsaneva-Atanasova, K. (2018). Relaxation oscillations and canards in the Jirsa-Kelso excitator model: global flow perspective. *Eur. Phys. J. Spec. Top.* 227, 591–601. doi: 10.1140/epjst/e2018-00129-2

Słowiński, P., Al-Ramadhani, S., and Tsaneva-Atanasova, K. (2020). Neurologically motivated coupling functions in models of motor coordination. *SIAM J. Appl. Dyn. Syst.* 19, 208–232. doi: 10.1137/19M1279381

Solfo, A., Luccio, R., and van Leeuwen, C. (2019). Sensorimotor coordination generates extended agency. *Cogn. Syst. Res.* 55, 219–244. doi: 10.1016/j.cogsys.2019.01.010

Soodak, H., and Iberall, A. (1978). Homeokinetics: a physical science for complex systems. *Science* 201, 579–582. doi: 10.1126/science.201.4356.579

Sperduti, M., Guionnet, S., Fossati, P., and Nadel, J. (2014). Mirror Neuron System and Mentalizing System connect during online social interaction. *Cogn. process.* 15, 307–316. doi: 10.1007/s10339-014-0600-x

Stewart, I. (1999). Designer differential equations for animal locomotion. *Complexity* 5, 12–22. doi: 10.1002/(sici)1099-0526(199911/12)5:2<12::aid-cplx3>3.0.co;2-m

Strogatz, S. H., and Stewart, I. (1993). Coupled oscillators and biological synchronization. *Sci. Am.* 269, 102–109. doi: 10.1038/scientificamerican1293-102

Taga, G., Yamaguchi, Y., and Shimizu, H. (1991). Self-organized control of bipedal locomotion by neural oscillators in unpredictable environment. *Biol. Cybern.* 65, 147–159. doi: 10.1007/bf00198086

Thelen, E., and Smith, L. B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press.

Tognoli, E. (2008). "EEG coordination dynamics: neuromarkers of social coordination," in *Coordination: Neural, Behavioral and Social Dynamics*, eds A. Fuchs and V. K. Jirsa (Berlin: Springer), 309–323. doi: 10.1007/978-3-540-74479-5_15

Tognoli, E., Dumas, G., and Kelso, J. A. S. (2018a). A roadmap to computational social neuroscience. *Cogn. Neurodyn.* 12, 135–140. doi: 10.1007/s11571-017-9462-0

Tognoli, E., Zhang, M., and Kelso, J. A. S. (2018b). "On the nature of coordination in nature," In *Advances in Cognitive Neurodynamics (VI)*, eds J. M. Delgado-García, X. Pan, R. Sánchez-Campusano, R. Wang (Singapore: Springer), 375–382.

Tognoli, E., and Kelso, J. A. S. (2009). Brain coordination dynamics: true and false faces of phase synchrony and metastability. *Prog. Neurobiol.* 87, 31–40. doi: 10.1016/j.pneurobio.2008.09.014

Tognoli, E., and Kelso, J. A. S. (2014a). Enlarging the scope: grasping brain complexity. *Front. Syst. Neurosci.* 8:122. doi: 10.3389/fnsys.2014.00122

Tognoli, E., and Kelso, J. A. S. (2014b). The metastable brain. *Neuron* 81, 35–48. doi: 10.1016/j.neuron.2013.12.022

Tognoli, E., and Kelso, J. A. S. (2015). The coordination dynamics of social neuromarkers. *Front. Hum. Neurosci.* 9:563. doi: 10.3389/fnhum.2015.00563

Tognoli, E., Lagarde, J., DeGuzman, G. C., and Kelso, J. A. S. (2007). The phi complex as a neuromarker of human social coordination. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8190–8195. doi: 10.1073/pnas.0611453104

Tschacher, W., Rees, G. M., and Ramseyer, F. (2014). Nonverbal synchrony and affect in dyadic interactions. *Front. Psychol.* 5:1323. doi: 10.3389/fpsyg.2014.01323

Tuller, B., and Kelso, J. A. S. (1989). Environmentally-specified patterns of movement coordination in normal and split-brain subjects. *Exp. Brain Res.* 75, 306–316.

Turrigiano, G., Abbott, L. F., and Marder, E. (1994). Activity-dependent changes in the intrinsic properties of cultured neurons. *Science* 264, 974–977. doi: 10.1126/science.8178157

Uhlhaas, P., Pipa, G., Lima, B., Melloni, L., Neuenschwander, S., Nikolić, D., et al. (2009). Neural synchrony in cortical networks: history, concept and current status. *Front. Integr. Neurosci.* 3:17. doi: 10.3389/neuro.07.017.2009

Van de Steen, F., Faes, L., Karahan, E., Songsiri, J., Valdes-Sosa, P. A., and Marinazzo, D. (2019). Critical comments on EEG sensor space dynamical connectivity analysis. *Brain Topogr.* 32, 643–654. doi: 10.1007/s10548-016-0538-7

van Wijk, B., Beek, P. J., and Daffertshofer, A. (2012). Neural synchrony within the motor system: what have we learned so far? *Front. Hum. Neurosci.* 6:252. doi: 10.3389/fnhum.2012.00252

Varela, F., Lachaux, J. P., Rodriguez, E., and Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. *Nat. Rev. Neurosci.* 2, 229–239. doi: 10.1038/35067550

von der Malsburg, C. (1995). Binding in models of perception and brain function. *Curr. Opin. Neurobiol* 5, 520–526. doi: 10.1016/0959-4388(95)80014-x

Wallenstein, G. V., Kelso, J. A. S., and Bressler, S. L. (1995). Phase transitions in spatiotemporal patterns of brain activity and behavior. *Phys. D* 84, 626–634. doi: 10.1016/0167-2789(95)00056-a

Walton, A. E., Richardson, M. J., Langland-Hassan, P., and Chemero, A. (2015). Improvisation and the self-organization of multiple musical bodies. *Front. psychol.* 6:313. doi: 10.3389/fpsyg.2015.00313

Winfree, A. T. (2001). *The geometry of Biological Time*, Vol. 12. Berlin: Springer Science & Business Media.

Yang, T., Chi, Y., Zhu, S., Gong, Y., and Jin, R. (2011). Detecting communities and their evolutions in dynamic social networks—a Bayesian approach. *Mach. Learn.* 82, 157–189. doi: 10.1007/s10994-010-5214-7

Yates, F. E. (1982). "Systems analysis of hormone action: principles and strategies," in *Biological Regulation and Development*, eds R. F. Goldberger and K. F. Yamamoto (New York, NY: Plenum), 23–97.

Yuste, R., MacLean, J. N., Smith, J., and Lansner, A. (2005). The cortex as a central pattern generator. *Nat. Rev. Neurosci.* 6, 477–483.

Zanone, P. G., and Kelso, J. A. S. (1992). Evolution of behavioral attractors with learning: nonequilibrium phase transitions. *J. Exp. Psychol.* 18, 403–421. doi: 10.1037/0096-1523.18.2.403

Zanone, P. G., and Kelso, J. A. S. (1997). Coordination dynamics of learning and transfer: collective and component levels. *J. Exp. Psychol.* 23, 1454–1480. doi: 10.1037/0096-1523.23.5.1454

Zhang, M., Beetle, C., Kelso, J. A. S., and Tognoli, E. (2019). Connecting empirical phenomena and theoretical models of biological coordination across scales. *J. R. Soc. Interface* 16:20190360. doi: 10.1098/rsif.2019.0360

Zhang, M., Dumas, G., Kelso, J. A. S., and Tognoli, E. (2016). Enhanced emotional responses during social coordination with a virtual partner. *Int. J. Psychophysiol.* 104, 33–43. doi: 10.1016/j.ijpsycho.2016.04.001

Zhang, M., Kalies, W. D., Kelso, J. A. S., and Tognoli, E. (2020). Topological portraits of multiscale coordination dynamics. *J. Neurosci. Methods* 339:108672. doi: 10.1016/j.jneumeth.2020.108672

Zhang, M., Kelso, J. A. S., and Tognoli, E. (2018). Critical diversity: divided or united states of social coordination. *PLoS One* 13:e0193843. doi: 10.1371/journal.pone.0193843

Zhou, G., Bourguignon, M., Parkkonen, L., and Hari, R. (2016). Neural signatures of hand kinematics in leaders vs. followers: a dual-MEG study. *Neuroimage* 125, 731–738. doi: 10.1016/j.neuroimage.2015.11.002

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Human Neuroscience

BRIEF RESEARCH REPORT
published: 04 September 2020
doi: 10.3389/fnhum.2020.560567

The Feeling Is Mutual: Clarity of Haptics-Mediated Social Perception Is Not Associated With the Recognition of the Other, Only With Recognition of Each Other

Tom Froese[1]*, *Leonardo Zapata-Fonseca*[2,3,4], *Iwin Leenen*[5] *and Ruben Fossion*[3,6]

[1]Embodied Cognitive Science Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan, [2]Plan of Combined Studies in Medicine (PECEM), Faculty of Medicine, National Autonomous University of Mexico, Mexico City, Mexico, [3]Center for the Sciences of Complexity (C3), National Autonomous University of Mexico, Mexico City, Mexico, [4]Section Phenomenological Psychopathology and Psychotherapy, Department of General Psychiatry, Center of Psychosocial Medicine, University of Heidelberg, Heidelberg, Germany, [5]Faculty of Psychology, National Autonomous University of Mexico, Mexico City, Mexico, [6]Institute of Nuclear Sciences, National Autonomous University of Mexico, Mexico City, Mexico

OPEN ACCESS

Edited by:
Andreas K. Engel,
University Medical Center
Hamburg-Eppendorf, Germany

Reviewed by:
Enrica Laura Santarcangelo,
University of Pisa, Italy
Eleonora Gentile,
University of Bari Aldo Moro, Italy

***Correspondence:**
Tom Froese
tom.froese@oist.jp

Specialty section:
This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

Received: *09 May 2020*
Accepted: *17 August 2020*
Published: *04 September 2020*

Citation:
Froese T, Zapata-Fonseca L, Leenen I and Fossion R (2020) The Feeling Is Mutual: Clarity of Haptics-Mediated Social Perception Is Not Associated With the Recognition of the Other, Only With Recognition of Each Other. Front. Hum. Neurosci. 14:560567. doi: 10.3389/fnhum.2020.560567

The enactive theory of perception hypothesizes that perceptual access to objects depends on the mastery of sensorimotor contingencies, that is, on the know-how of the regular ways in which changes in sensations depend on changes in movements. This hypothesis can be extended into the social domain: perception of other minds is constituted by mastery of self-other contingencies, that is, by the know-how of the regular ways in which changes in others' movements depend on changes in one's movements. We investigated this proposal using the perceptual crossing paradigm, in which pairs of players are required to locate each other in an invisible one-dimensional virtual space by using a minimal haptic interface. We recorded and analyzed the real-time embodied social interaction of 10 pairs of adult participants. The results reveal a process of implicit perceptual learning: on average, clarity of perceiving the other's presence increased over trials and then stabilized. However, a clearer perception of the other was not associated with correctness of recognition as such, but with both players correctly recognizing each other. Furthermore, the moments of correct mutual recognition tended to happen within seconds. The fact that changes in social experience can only be explained by the successful performance at the level of the dyad, and that this veridical mutual perception tends toward synchronization, lead us to hypothesize that integration of neural activity across both players played a role.

Keywords: embodied cognition, social cognition, enactive approach, virtual reality, agency detection, perceptual awareness scale

INTRODUCTION

Imagine you are going on a romantic date at the cinema. Inside the movie theater you sit down next to your date, but it is so dark that you cannot see each other, leaving you uncertain about their presence. At some point during the movie you feel your date's hand touching your hand, and you start holding hands. Your experience of watching the movie is transformed, as it takes on a more distinctively social quality: "we" are sharing this experience with each other. How can such embodied social interaction have this profound effect on an individual's perceptual experience?

This study aimed to investigate how a person's real-time tactile interaction with other people can make an irreducible difference to how that person experiences their self, others, and the world, an effect which has been referred to as "genuine intersubjectivity" (Froese, 2018; Froese and Krueger, forthcoming). It builds on the key role of interpersonal contingencies that emerge from the coupling of human bodies Dumas et al., 2014). This aim and basis stand in sharp contrast to what has been characterized as the "methodological individualism" of traditional cognitive science (Boden, 2006), which in its more extreme formulations has even taken an isolated brain as the in-principle sufficient basis of all social experience (Searle, 1990). Nevertheless, the notion of genuine intersubjectivity is consistent with a small but growing number of psychological and neuroscientific experiments as well as agent-based simulation studies, which point to the constitutive role of social interaction for social cognition (e.g., De Jaegher et al., 2010; Schilbach et al., 2013; Candadai et al., 2019).

A particularly promising methodology for studying the effects of real-time embodied social interaction is the so-called "perceptual crossing" paradigm, which was originally proposed by Lenay and colleagues (Lenay et al., 2006; Auvray et al., 2009). Pairs of participants are connected to an invisible 1D virtual space using a minimal haptic computer interface, and their task is to locate the other person's avatar based on the patterns of interaction while avoiding distractor objects (see **Figure 1**). This paradigm has inspired several experimental variations and different applications (for a review, see Auvray and Rohde, 2012; Deschamps et al., 2016; Zapata-Fonseca et al., 2018; Barone et al., 2020). Since the first studies, there have been discussions of participants' anecdotal reports of their social encounters (Lenay, 2010; Auvray and Rohde, 2012). However, to investigate genuine intersubjectivity more experimentally, such studies must also include an explicit assessment of participants' lived experience.

An important step in this direction was provided by a variation of the perceptual crossing paradigm by Froese et al. (2014a), in which participants were asked to rate the clarity of their perception of the other's presence. They found that a participant's perceptual clarity was a joint achievement: it was not associated with a participant's correct identification of the other *per se*, but rather with bidirectional interactions that permitted both participants to successfully recognize each other. This conclusion was further supported by subsequent explorations of the same dataset (Froese et al., 2014b; Zapata-Fonseca et al., 2016; Kojima et al., 2017). This was the first empirical proof of the concept of genuine intersubjectivity.

However, given that this was an isolated experiment, it remained to be seen whether it could be replicated. This has now been accomplished by Hermans et al. (2020), as well as by the present study. Hermans et al. (2020) implemented a shorter variation of Froese et al.'s (2014a) perceptual crossing paradigm as part of a longitudinal population study of adolescents. Importantly, despite this reduction in overall interaction time, as well as differences in the target population and in the assessment of participant's experience, they also found that subjective experience was highest specifically in trials when both participants were jointly correct in detecting each other's presence.

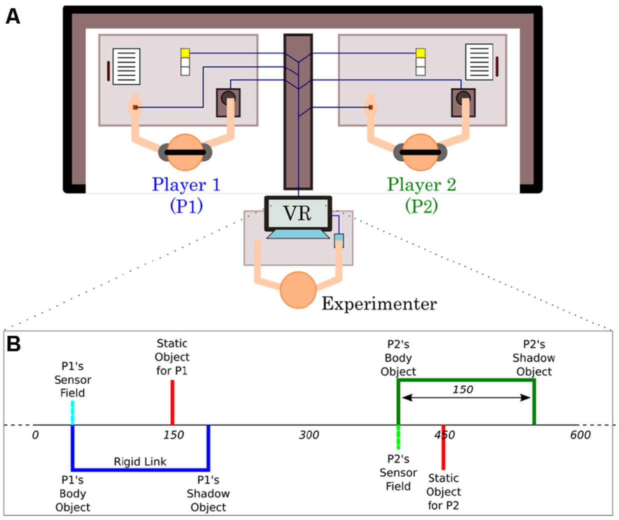


FIGURE 1 | Experimental setup of the perceptual crossing paradigm. **(A)** The physical setup. The two participants can only engage with each other *via* a haptic human-computer interface that reduces their scope for bodily interaction to a minimum of horizontal left to right movement and tactile sensation. Each player's interface consists of two parts: (1) a trackball that controls the displacement of their "avatar" in an invisible 1D virtual environment; and (2) a hand-held haptic feedback device that vibrates at a constant frequency for as long as the avatar overlaps another virtual object and remains off otherwise. **(B)** The virtual setup. Players are embodied as minimal avatars on an invisible line that wraps around after 600 units of space. Each avatar consists of a binary contact sensor and a body object. Unbeknownst to the players a "shadow" object is attached to each avatar body at a fixed distance of 150 units. There are also two static objects, one for each player. All objects are four units long and can therefore only be distinguished interactively in terms of their different affordances for engagement.

Hypotheses

In the present study, we replicated Froese et al.'s experimental setup and tested several hypotheses related to genuine intersubjectivity. This concept is best approached by drawing on the theoretical resources of the phenomenological tradition in philosophy, especially its work on the phenomenology of direct social perception (Gallagher, 2008; Krueger, 2012), and of the enactive approach to cognitive science, especially the enactive theory of perception (Noë, 2012) and its extension to social perception (De Jaegher, 2009; McGann and De Jaegher, 2009). In essence, this theory holds that object perception consists in the skillful regulation of sensorimotor interaction, which involves knowing how sensations of the object would change concerning possible movements of one's body, and which provides direct access to the object of perception. An individual's mastery of these dependencies of sensations on bodily movements, that is, of so-called sensorimotor contingencies, entails a better perceptual grasp of the perceived object.

The same idea of sensorimotor contingencies can also be applied to the perception of another person. It is known that the motor system is involved in social cognition, which is often interpreted as simulation or mirroring (Rizzolatti and Sinigaglia, 2016), but can also be considered as enactive perception in a social context (Gallagher, 2009; but see Gallese, 2017). More specifically, in line with sensorimotor theory (O'Regan et al., 2005), it has been proposed that social perception consists in the skillful co-regulation of social interaction (De Jaegher et al., 2010; Froese and Di Paolo, 2011), which involves both participants knowing how sensations caused by the other's bodily movements would change concerning their possible bodily movements, and the mastery of these "self-other contingencies" thereby provides access to the other person (McGann and De Jaegher, 2009).

According to this enactive theory, there is a crucial difference between social perception and object perception in terms of their respective conditions of successful perceptual access: both depend on skillful regulation of interaction, but in the case of perceiving another person as another person this access additionally depends on a complementary skillful response by the other person. If the other does not respond appropriately, the perceptual situation would be more akin to that of object perception, for example of perceiving the other's physical body. This theory, therefore, predicts that the basis of social perception goes beyond the individual perceiver to include another perceiver. However, it is not yet entirely clear what this inter-personal basis of social perception means for the perceiver's experience. At least two possibilities present themselves, which we will refer to as weak and strong forms of genuine intersubjectivity, respectively:

(1) *Weak genuine intersubjectivity*: This possibility accepts that the basis of social perception can be distributed across two persons, such that one person's perception of the other person is partly constituted by their ongoing social interaction. However, it remains conservative about the boundaries of consciousness, because it still maintains that each person's experience remains the property of only that individual. This implies that there are two independent, non-overlapping experiences; each person's social perception can be shaped by the other's movements, but without ever constituting a single, jointly shared moment of experience.
(2) *Strong genuine intersubjectivity*: This possibility also accepts that the basis of social perception can be distributed across two persons, but it is more liberal about the boundaries of consciousness. It rejects the claim that two interacting persons must always have two independent experiences and instead accepts the possibility that an interactively extended basis can also give rise to one jointly unfolding experience, for instance of mutually perceiving each other. This implies that this single, jointly shared experience is better characterized as a moment of co-presence that is grasped from each perceiver's specific point of view.

The possibility of weak genuine intersubjectivity implies that two people in interaction can have their experience shaped by that ongoing interaction at different times. The possibility of strong genuine intersubjectivity, on the other hand, implies tighter inter-personal integration, which fits well with growing evidence that there is a synchronization of neural activity across brains during social interaction, including in the faster frequency bands, and that this is the basis for inter-personal neuronal integration that has implications for different aspects of social cognition (Valencia and Froese, 2020).

More specifically, Froese (2018) has suggested extending Varela's (1999) neuro-phenomenological analysis of present-time consciousness to the social domain. The idea that genuine intersubjectivity requires interpersonal integration at the most fundamental level of temporality resonates with research in the phenomenology of consciousness (Rodemeyer, 2010), and complements it with a scientific methodology. Varela highlights that the conscious moment of "now" is not an instant but has a duration of 1–3 s. He argues that this duration results from the amount of time it takes for neural activity in an individual's brain to become transiently integrated *via* long-range synchronization.

Accordingly, an attractive hypothesis of how two people could experience that "we" are sharing one and the same "now," is that their co-regulated social interaction caused their neural activity to become synchronized. We did not directly investigate neural activity in the current experiment, but if this hypothesis is on the right track, then we would expect the time scale of 3 s to be relevant for mutual veridical perception.

We set out to investigate this theory of social perception and the possibilities of weak and strong genuine intersubjectivity in terms of the following hypotheses:

(1) We hypothesized that participants' capacity to correctly recognize the other will increase over trials, as they learn to redeploy their existing skill of embodied social interaction *via* the haptic computer interface.
(2) We hypothesized that clarity of the other's perceived presence will increase over trials, as perceptual learning will improve access to the other person.
(3) We hypothesized that increased clarity of the other's presence will not be explained by the correctness of an individual perceiver's recognition of the other person, but by the correctness of both perceivers' recognition of each other, as the shift from object perception to social perception involves a shift from regulation to co-regulation that offers a shared opportunity of recognition to both participants.
(4) We hypothesized that the moments of recognition in trials where both participants correctly recognize each other will tend to be synchronized, specifically in the time scale of 3 s, as a reflection of the shift from individual action to joint action.
(5) We hypothesized that the synchronization of moments of recognition will correlate with the clarity of the perception of the other person, specifically in the time scale of 3 s, as a reflection of the shift from two individual experiences to one intersubjectively shared experience.

Hypotheses 1 and 2 are intended to demonstrate, in behavioral and experiential terms respectively, that an embodied skill of perceiving others is being (re-)acquired throughout the experiment. Hypothesis 3 is aimed at supporting the concept of weak genuine intersubjectivity, i.e., the idea that co-regulated interpersonal interaction makes an irreducible difference to individual experience. Hypothesis 4 and 5 are aimed at supporting the concept of strong genuine intersubjectivity, i.e., that there is a single joint action giving rise to one shared moment of veridical mutual recognition that "we" are now experiencing, based on the assumption that this fusion of individual streams of experiencing will require integration of neural activity across both participants in the time scale of seconds.

MATERIALS AND METHODS

We employed a version of the perceptual crossing experiment that was designed to capture subjective reports of participants' perceptual awareness of the other's presence to investigate its sensorimotor basis (Froese et al., 2014a,b; Kojima et al., 2017).

Experimental Equipment

In the perceptual crossing experiment, two participants are seated at separate desks such that they cannot perceive each other visually; they also wear noise-canceling headphones to prevent mutual auditory perception (**Figure 1A**). Each participant can establish contact with her partner only *via* a simple human-computer interface consisting of two components: (1) a trackball for making horizontal movements; and (2) a hand-located vibration motor that is either on or off. The trackball is operated with the dominant hand and it controls the motions of an avatar located in an invisible circular 1D virtual space (**Figure 1B**). The motor vibrates as long as the avatar overlaps with another virtual object in this space. Each participant can encounter three objects: (i) the other's avatar; (ii) a moving object that "shadows" the other's avatar by following the same trajectory at a distance; and (iii) a static object. Regardless of the object type, the vibratory feedback is only on (and off otherwise).

Experimental Procedure

Participants were told to work as a team and were asked to come up with a team name, with which they would be ranked against other teams participating in the study. They were instructed to navigate through the invisible shared space to find each other. They were asked to signal with a click (only once per trial) when they became aware of interacting with the other player; for each click correctly identifying their partner the team would gain 1 point, for each wrong click they would lose 1 point. However, no feedback about the correctness of clicks was provided during the experiment. Participants were first individually familiarized with the human-computer interface. Then each pair was tested for 20 trials, consisting of 60 s each (due to errors the last five trials and the last trial of teams 1 and 6, respectively, were not recorded). This is the first time the perceptual crossing experiment has been run for 20 trials; Froese et al. (2014a) had employed 15 trials, whereas Hermans et al. (2020) used only six trials. The aim of using 20 trials was to see if there would be a change in results if participants have more time to interact.

After each trial, the experience of the players was evaluated through questionnaires based on a version of the Perceptual Awareness Scale (PAS), which in its original formulation was used for visual perception (Ramsøy and Overgaard, 2004), but was adapted by Froese et al. (2014a) for social perception. After each trial in which a participant clicked, his or her awareness of the other's presence at the moment of the click was assessed with values from one to four corresponding to "no experience," "ambiguous experience," "almost clear experience," and "clear experience," respectively.

The target of each click was categorized as one of the three objects or as unknown. Target assignment was first done automatically based on calculated distances at the time of the click, following (Froese et al., 2014a). This was followed by trial-by-trial visual inspection of plots of the movement trajectories by TF and LZF (see **Supplementary Trial Data**). Any discrepancies in the target assignment were resolved by TF and LZF in discussion with IL.

Participants

Twenty adults took part in the perceptual crossing experiment, ages ranged between 18 and 47 years old (median of age = 28), and there were 6 women and 14 men (for details, see Appendix of the **Supplementary Statistical Information**). Participants were recruited from acquaintances at UNAM in Mexico City. Only healthy volunteers were considered; individuals with neurological, psychiatric, or movement disorders (clinically diagnosed) were excluded. Participation was voluntary and all participants gave their informed consent. Ten teams were created as pairs of volunteers became available. Three teams were composed of strangers, while participants in other teams had some history of interaction from before and some were friends. Some participants were familiar with perceptual crossing from the literature, but none had previous experience of participating in such experiments. Data collection took place between April and May 2018.

Statistical Analyses

We specified a statistical model that allowed us to simultaneously examine the relations among the following variables: (a) individual success and (b) joint success in recognizing the other player, (c) the PAS-responses, (d) the inter-click delay (dichotomized, with a value of 1 in case that both players clicked within the same timeframe of 3 s, and 0 otherwise) and (e) the trial number, by which we modeled a learning process for individual successes and PAS-responses across trials. The effects were specified as depicted in **Figure 2**, by including (binary or ordinal) probit regression submodels for the endogenous variables (individual success, joint success, and PAS-responses); the learning process was modeled through piece-wise regression, with first a learning stage allowing improvement, and subsequently, a consolidation stage where the individual is assumed to operate at the same level. The model is hierarchical (i.e., it includes random

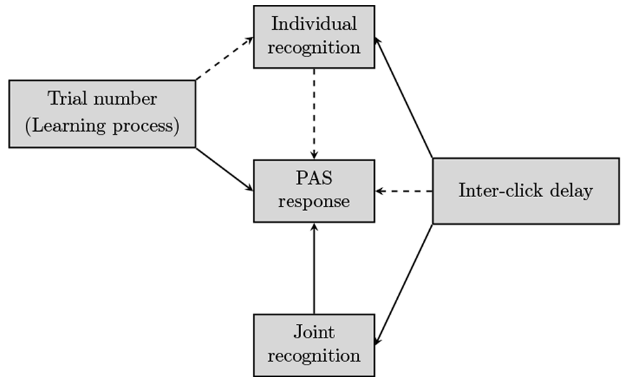


FIGURE 2 | Relations among observed variables at the level of the team (joint recognition, inter-click delay), trial (trial number), and individual participant [individual recognition, Perceptual Awareness Scale (PAS) response], as modeled in the statistical analysis. Dotted arrows indicate that for those relations no direct effect was found.

effects) as it takes into account the nested structure of the data (with both individuals responding to each trial, and 20 trials for each team) and was fitted using a fully Bayesian approach.

All data can be found in the **Supplementary Data Table**, while further details of the analysis (together with a complete description of the results) can be found in the **Supplementary Statistical Information**.

RESULTS

In total, we recorded 194 trials, for a possibility of 194 × 2 = 388 clicks and PAS-responses. A click was produced in 311 of these 388 cases (i.e., 80%). Of these 311 clicks, 225 (72%) were correct clicks, responding to the other's avatar. In 79 (41%) of the 194 trials, players produced jointly correct clicks. In 48 (25%) of the 194 trials, both players clicked but either or both gave an incorrect click. There were 307 PAS-responses (four clicks did not receive a PAS-response).

Overall, the frequencies for the four PAS-response categories, ordered from "no experience" to "clear experience," are 20 (6% out of 307), 91 (30%), 100 (33%), and 96 (31%), respectively. For the 85 PAS-responses associated with an *incorrect* click, the frequencies are 6 (7%), 31 (37%), 34 (40%), 14 (16%), respectively. Responses conditional upon a (65) *correct* click in non-jointly successful trials were: 6 (9%), 30 (46%), 16 (25%), and 13 (20%), while responses conditional upon a (157) *correct* click in *jointly* successful trials were: 8 (5%), 30 (19%), 50 (32%), and 69 (44%).

We summarize here the main findings from the detailed statistical analysis included as **Supplementary Statistical Information**, focusing on the hypotheses listed in the introduction:

(1) We used a piecewise regression model for the individual learning process of correctly recognizing the other player, with a learning stage and a consolidation stage. On average, learning takes place between trial 1 and trial 3.2 [95%-high posterior density interval = (1.7, 4.9)]; however, there are large differences, with individuals who apparently do not enter in a learning process and others for whom learning takes place until half the experiment.
(2) Concerning the learning process on experiential clarity, the breakpoint that separates the learning and consolidation stage is situated, on average, at trial 5.1 [95%-HPDI = (1.1, 9.1)], but again with a relatively large variance[1].
(3) The results do not show evidence of an effect of individual success in recognizing the other on the PAS-responses [with an estimated effect on the probit scale of $\beta = -0.16$; 95%-HPDI = (−0.55, 0.20)], whereas joint success does lead to higher PAS-values [$\beta = 0.69$; 95%-HPDI = (0.31, 1.06)].
(4) A short inter-click delay, of less than 3 s, goes with a higher probability of individual success [with an estimated effect on the probit scale of $\beta = 0.70$, 95%-HPDI = (0.16, 1.29)] as well as a higher probability of joint success in recognizing the other [$\beta = 1.01$, 95%-HPDI = (0.40, 1.59)].
(5) There is no clear evidence of a direct association between short inter-click delays and PAS-responses [$\beta = 0.14$, 95%-HPDI = (−0.28, 0.54)]. Note, however, that there is an indirect effect given that short inter-click delays are associated with higher probabilities of joint successful recognition (see the previous point), which in turn leads to higher PAS-responses (see Point 3).

DISCUSSION

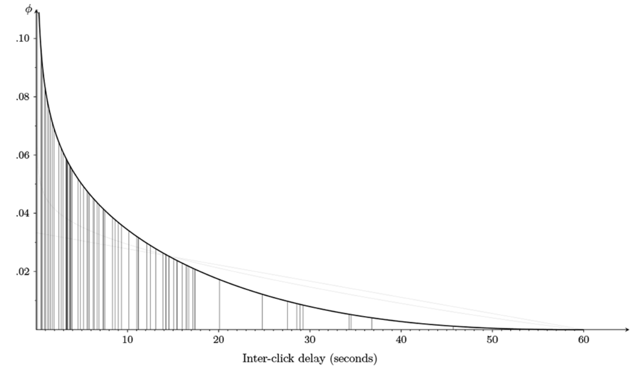
These results support genuine intersubjectivity, although evidence for strong genuine intersubjectivity remains indirect.

First, on average, participants' perceptual experience of the other's presence became clearer during the experiment. This change in experience tended to stabilize within six trials, and our study thereby supports Hermans et al.'s (2020) decision to run a shorter experiment of six trials. However, it is noteworthy that we did not find compelling evidence that, on average, participants improved their capacity to click correctly. Some improved quickly, others were slow learners, and some never improved even though we had extended the number of trials to 20 trials. Future work could investigate the reasons for this diversity in learning outcomes; presumably, the recognition task is facilitated if participants have a history of close interaction, and there may also be an influence of individuals' age and sex, which are factors that we did not take into consideration in the current analysis. Yet the fact that clarity of social presence increased, and did so independently of individuals' success at objectively recognizing the other, also suggests that an explanation of this phenomenological change should look beyond individuals.

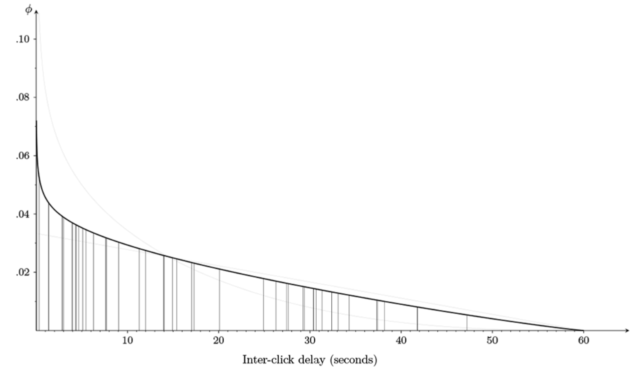
Second, indeed, we found compelling evidence for genuine intersubjectivity: an individual's increased perceptual clarity of the other's presence could only be explained by taking into account task performance at the level of the dyad, and not at the level of the individuals. In other words, an individual's correct

---

[1] As explained in the **Supplementary Statistical Information** (and due to the small sample size), this and the previous result must be interpreted with caution, as the associated parameter estimates have wide credibility intervals.



Joint-success trials



Non-joint-success trials

FIGURE 3 | Estimated distribution of inter-click delays for joint-success trials and non-joint-success trials (i.e., where either one or both players produced a click that did not correctly identify the other). The null hypothesis (based on independent uniform distributions of both players' clicks) is represented by a thin gray line. We graphed the estimated density function separately for both conditions: joint-success trials (upper figure) and non-joint-success trials (lower figure). To allow a better comparison, we show (by a thin gray curve) the estimated density function of the other condition in both figures. Furthermore, the vertical lines connecting the density with the abscissa represent the observed values (79 inter-click intervals for joint success; 48 for non-joint success). Shorter inter-click intervals are more likely than longer intervals. Importantly, comparing joint-success to non-joint-success trials, inter-click intervals tend to be even shorter, with for example the estimated probability that an interval is below 3 s being equal to 0.25 and 0.14, respectively.

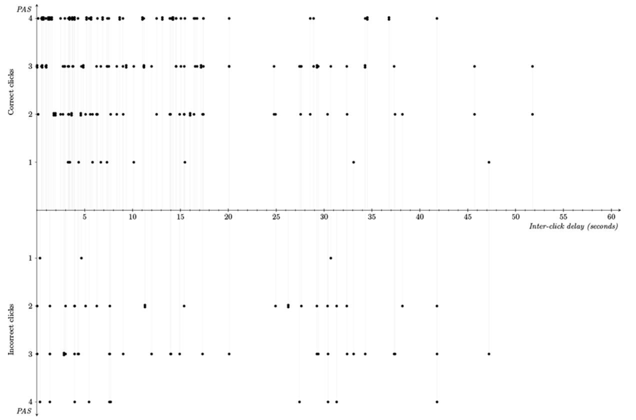


FIGURE 4 | Graphical representation of the relationship between individuals' responses. Responses were associated with the PAS, individuals' correctly recognizing the other player (i.e., correct or incorrect click), and the dyads' inter-click delay in trials where both members of the dyad produced a click. The thin vertical gray lines connect both responses in a dyad that correspond to the same trial and projects the corresponding points to the horizontal axis to show the corresponding inter-click delay.

recognition of the other was necessary but not sufficient to explain the changed quality of perceptual experience; the other's involvement, as measured in terms of their correct recognition of the self, was also necessary.

It is reassuring that we managed to replicate this dependence of perceptual experience on the interaction process using a comprehensive statistical model. This key finding is in tension with the traditional view of perception as a brain-based process of furnishing mental representations inside of the individual. Instead, it fits better with the enactive view that the basis of perceptual experience extends into sensorimotor interaction, which in the case of social perception also includes a relationship with another subject. Future work could try to uncover these distinctive movement patterns. Based on past analyses, we can expect increases in interdependence as measured by, for example, turn-taking, cross-correlations, and transfer entropy in sensorimotor dynamics (Kojima et al., 2017).

Third, we found indications of the possibility of interpersonally extended experience of veridical mutual recognition (strong genuine intersubjectivity). As revealed by **Figure 3**, comparing jointly to non-jointly correct clicks, inter-click intervals tended to be shorter, with the estimated probability that an interval is below 3 s being equal to 0.25 and 0.14, respectively. Also, there was an indirect effect of inter-click intervals within 3 s on increased clarity of social presence, given that those intervals are associated with a higher probability of joint success, which is associated with higher PAS ratings. This tendency of jointly correct 3-s inter-click intervals to be associated with higher PAS ratings can also be seen in **Figure 4**; future studies with larger sample sizes may still uncover a direct effect.

On the other hand, perhaps we did not develop our hypotheses carefully enough. In contrast to the representational approach to perception, the enactive approach has emphasized that perceptual experience is constituted by certain kinds of organism-environment interaction; hence, the experience is not an entity that is somehow separate from that interaction. Or, more strongly, the perceptual experience is identical with the sensorimotor interaction (Myin and Zahnoun, 2018), or at least not something added to the process (Froese and Taguchi, 2019). Accordingly, the fact that the effect on the clarity of social perception is specifically mediated by jointly correct recognition is consistent with another intriguing possibility: the social quality of experience is constituted by their social interaction. This would also account for the puzzling finding that the residual correlation between PAS-responses of both individuals in a dyad is close to zero, which implies that the other effects included in the model (the learning process, individual and joint success in recognizing the other, and the inter-click delay) may fully account for this correlation.

Overall, this is a promising line of investigation for future work, and it should be possible to further clarify the basis and extent of genuine intersubjectivity by increasing the sample size and by applying a hyperscanning approach to measure neural activity of both participants.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Faculty of Medicine, National Autonomous University of Mexico. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

TF and LZ-F conceived the idea for this study and designed the experiment. LZ-F and RF carried out the experiment. TF and IL analyzed the data. TF and LZ-F interpreted the results. TF drafted the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

LZ-F's contribution was supported by scholarships from Mexico's National Council of Science and Technology (Consejo Nacional de Ciencia y Tecnología; CONACyT, #638215) and the German Academic Exchange Service (Deutscher Akademischer Austauschdienst; DAAD, #91623295). RF and LZ-F acknowledge support from UNAM-DGAPA-PAPIIT project #IA105017 (Dirección General de Asuntos del Personal Académico, Universidad Nacional Autónoma de México) and CONACyT project #167441. RF acknowledges support from the Newton Advanced Fellowship of the Academy of Medical Sciences.

ACKNOWLEDGMENTS

We acknowledge the support of Ximena Gonzalez-Grandon with the data collection and José de Jesús Naveja with the programming of automatic click target assignments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2020.560567/full#supplementary-material.

REFERENCES

Auvray, M., and Rohde, M. (2012). Perceptual crossing: the simplest online paradigm. *Front. Hum. Neurosci.* 6:181. doi: 10.3389/fnhum.2012.00181

Auvray, M., Lenay, C., and Stewart, J. (2009). Perceptual interactions in a minimalist virtual environment. *New Ideas Psychol.* 27, 32–47. doi: 10.1016/j.newideapsych.2007.12.002

Barone, P., Bedia, M. G., and Gomila, A. (2020). A minimal turing test: reciprocal sensorimotor contingencies for interaction detection. *Front. Hum. Neurosci.* 14:102. doi: 10.3389/fnhum.2020.00102

Boden, M. A. (2006). Of islands and interactions. *J. Conscious. Stud.* 13, 53–63.

Candadai, M., Setzler, M., Izquierdo, E. J., and Froese, T. (2019). Embodied dyadic interaction increases complexity of neural dynamics: a minimal agent-based simulation model. *Front. Psychol.* 10:540. doi: 10.3389/fpsyg.2019.00540

De Jaegher, H. (2009). Social understanding through direct perception? yes, by interacting. *Conscious. Cogn.* 18, 535–542. doi: 10.1016/j.concog.2008.10.007

De Jaegher, H., Di Paolo, E., and Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends Cogn. Sci.* 14, 441–447. doi: 10.1016/j.tics.2010.06.009

Deschamps, L., Lenay, C., Rovira, K., Le Bihan, G., and Aubert, D. (2016). Joint perception of a shared object: a minimalist perceptual crossing experiment. *Front. Psychol.* 7:1059. doi: 10.3389/fpsyg.2016.01059

Dumas, G., Laroche, J., and Lehmann, A. (2014). Your body, my body, our coupling moves our bodies. *Front. Hum. Neurosci.* 8:1004. doi: 10.3389/fnhum.2014.01004

Froese, T. (2018). "Searching for the conditions of genuine intersubjectivity: from agent-based models to perceptual crossing experiments," in *The Oxford Handbook of 4E Cognition*, eds A. Newen, L. De Bruin and S. Gallagher (Oxford, UK: Oxford University Press), 163–186.

Froese, T., and Di Paolo, E. A. (2011). The enactive approach: theoretical sketches from cell to society. *Pragmat. Cogn.* 19, 1–36. doi: 10.1075/pc.19.1.01fro

Froese, T., and Krueger, J. (forthcoming). "Lost in the socially extended mind: Genuine intersubjectivity and disturbed self-other demarcation in schizophrenia," in *Time and Body: Phenomenological and Psychopathological Approaches*, eds C. Tewes and G. Stanghellini (Cambridge, UK: Cambridge University Press).

Froese, T., and Taguchi, S. (2019). The problem of meaning in AI and robotics: still with us after all these years. *Philos.* 4:14. doi: 10.3390/philosophies4020014

Froese, T., Iizuka, H., and Ikegami, T. (2014a). Embodied social interaction constitutes social cognition in pairs of humans: a minimalist virtual reality experiment. *Sci. Rep.* 4:3672. doi: 10.1038/srep03672

Froese, T., Iizuka, H., and Ikegami, T. (2014b). Using minimal human-computer interfaces for studying the interactive development of social awareness. *Front. Psychol.* 5:1061. doi: 10.3389/fpsyg.2014.01061

Gallagher, S. (2008). Direct perception in the intersubjective context. *Conscious. Cogn.* 17, 535–543. doi: 10.1016/j.concog.2008.03.003

Gallagher, S. (2009). "Neural simulation and social cognition," in *Mirror Neuron Systems*, ed. J. A. Pineda (New York, NY: Humana Press), 355–371.

Gallese, V. (2017). Neoteny and social cognition: a neuroscientific perspective on embodiment. in *Embodiment, Enaction and Culture: Investigating the Constitution of the Shared World*, eds C. Durt, T. Fuchs and C. Tewes (Cambridge, MA: MIT Press), 309–331.

Hermans, K. S. F. M., Kasanova, Z., Zapata-Fonseca, L., Lafit, G., Fossion, R., Froese, T., et al. (2020). Investigating real-time social interaction in pairs of adolescents with the perceptual crossing experiment. *Behav. Res. Methods.* doi: 10.3758/s13428-020-01378-4

Kojima, H., Froese, T., Oka, M., Iizuka, H., and Ikegami, T. (2017). A sensorimotor signature of the transition to conscious social perception: co-regulation of active and passive touch. *Front. Psychol.* 8:1778. doi: 10.3389/fpsyg.2017.01778

Krueger, J. (2012). Seeing mind in action. *Phenom. Cogn. Sci.* 11, 149–173. doi: 10.1007/s11097-011-9226-y

Lenay, C. (2010). "It's so touching": emotional value in distal contact. *Int. J. Des.* 4, 15–25.

Lenay, C., Auvray, M., Sebbah, F.-D., and Stewart, J. (2006). "Perception of an intentional subject: an enactive approach," in *Proceedings of the Third International Conference on Enactive Interfaces*, eds B. G. Bardy and D. Mottet (Montpellier: University Montpellier-1), 37–38.

McGann, M., and De Jaegher, H. (2009). Self-other contingencies: enacting social perception. *Phenom. Cogn. Sci.* 8, 417–437. doi: 10.1007/s11097-009-9141-7

Myin, E., and Zahnoun, F. (2018). Reincarnating the identity theory. *Front. Psychol.* 9:2044. doi: 10.3389/fpsyg.2018.02044

Noë, A. (2012). *Varieties of Presence*. Cambridge, MA: Harvard University Press.

O'Regan, J. K., Myin, E., and Noë, A. (2005). Sensory consciousness explained (better) in terms of "corporality" and "alerting capacity". *Phenom. Cogn. Sci.* 4, 369–387. doi: 10.1007/s11097-005-9000-0

Ramsøy, T. Z., and Overgaard, M. (2004). Introspection and subliminal perception. *Phenom. Cogn. Sci.* 3, 1–23. doi: 10.1023/b:phen.0000041900.30172.e8

Rizzolatti, G., and Sinigaglia, C. (2016). The mirror mechanism: a basic principle of brain function. *Nat. Rev. Neurosci.* 17, 757–765. doi: 10.1038/nrn.2016.135

Rodemeyer, L. M. (2010). *Intersubjective Temporality: It's About Time*. Dordrecht, The Netherlands: Springer.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., et al. (2013). Toward a second-person neuroscience. *Behav. Brain Sci.* 36, 393–414. doi: 10.1017/S0140525X12000660

Searle, J. R. (1990). "Collective intentions and actions," in *Intentions in Communication*, eds P. Cohen, J. Morgan and M. E. Pollack (Cambridge, MA: MIT Press), 401–416

Valencia, A. L., and Froese, T. (2020). What binds us? inter-brain neural synchronization and its implications for theories of human consciousness. *Neurosci. Conscious.* 2020:niaa010. doi: 10.1093/nc/niaa010

Varela, F. J. (1999). Present-time consciousness. *J. Conscious. Stud.* 6, 111–140.

Zapata-Fonseca, L., Dotov, D., Fossion, R., and Froese, T. (2016). Time-series analysis of embodied interaction: movement variability and complexity matching as dyadic properties. *Front. Psychol.* 7:1940. doi: 10.3389/fpsyg.2016.01940

Zapata-Fonseca, L., Froese, T., Schilbach, L., Vogeley, K., and Timmermans, B. (2018). Sensitivity to social contingency in adults with high-functioning autism during computer-mediated embodied interaction. *Behav. Sci.* 8:22. doi: 10.3390/bs8020022

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

OPINION
published: 10 September 2020
doi: 10.3389/fnhum.2020.00337

(New) Realist Social Cognition

Nicolás Araneda Hinrichs [1,2]*

[1] Laboratorio de Psicolingüística, Universidad de Concepción, Concepción, Chile, [2] Institut für Angewandte Linguistik und Translatologie, Universität Leipzig, Leipzig, Germany

Keywords: social cognition, sensorimotor coupling, new realism, ontology, epistemology, 4E-cognition

Edited by:
Lutz Jäncke,
University of Zurich, Switzerland

Reviewed by:
Marianna Delussi,
University of Bari Aldo Moro, Italy

****Correspondence:***
Nicolás Araneda Hinrichs
nqo@pm.me

Specialty section:
This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

Received: *30 April 2020*
Accepted: *31 July 2020*
Published: *10 September 2020*

Citation:
Araneda Hinrichs N (2020) (New)
Realist Social Cognition.
Front. Hum. Neurosci. 14:337.
doi: 10.3389/fnhum.2020.00337

INTRODUCTION

Low-level descriptions of interaction dynamics have been canonically approached by cognitive neuroscience through a representation-oriented and inference-based perspective, leading to a stable paradigmatic *plateau,* that no longer allows further construction of a completely coherent semiotic framework capable of accounting for currently unobserved characteristics of social cognition, which is forcibly situated and mostly occurs in interaction. Social contexts are saturated with information that remains invisibilized because of the use of mutually incommensurable conceptual metaphors throughout contemporary scientific discursive practices, despite the embodied turn led by 4E Cognition. A new turn toward realist ontology and epistemology is thus rendered as necessary to inform the gaps within cognitive neuroscience and ground its currently unfulfilled interdisciplinarity. Examples are drawn from research on language to make the case for each argument.

Trending cognitive neuroscience performs low-level descriptions of individual or group interactions by the use of state of the art techniques and methodologies. These observations can be defined as being close to the material niveau of the structure and functioning of our organism as a biological entity. Conscious processes like states of emotion, perception or belief formation –all of which motivate human behavior– transcend the reach of this scope, nonetheless. Thus, the general claims about these epistemic constructs, as a whole, should be more qualified.

For instance, psycholinguistics used to be quite English based and postulated general principles of which later turned to be proven that they were not replicable with other languages. Psychophysical cues in language processing need to be redefined epistemologically from a new materialistic perspective, in order to account for group learning and social transmission of knowledge. Theory of embodiment has provided the starting point for such a cultural approach of cognition, as it has been summarized by Storch and Tschacher (2014): “The social environment affects the embodied mind” (“Die soziale Umwelt beeinflusst den verkörperten Geist”).

KNOWING KNOWLEDGE

Cognitive Canon

Traditional cognitive theory is based mainly around symbolic manipulation, a process that consists of an analysis of incoming information that is fed into a processor, a unit that produces an outgoing result. Operations that mediate that process behave according to syntactic rules, through which this processor finds a semantic meaning.

This point of view, used as a model in linguistic research, has been widely spread, generating as a by-product its usage as a knowledge model of the human body in a generalized fashion (e.g., sensory information considered as input data), regarding ontogenic development and interaction. As has been ascertained so far (Barsalou, 1999; Meteyard et al., 2012; Urrutia and de Vega, 2012), physiology that supports language processing in human beings resists itself to be reduced to the notion of a mere processor; this is a major challenge for the development of artificial intelligence, deep learning and brain-to-computer interfaces. The canonic focus of symbolism, in its analysis, on the processor in regards to its structure –assuming entities are alike at an internal level–, has downplayed the phenomenal content created by the relation of an entity with their rather external counterpart.

Completely detached and exclusively extrospective perspectives have been rendered banal since the emergence of situated and intersubjective based theories, such as 4E-Cognition (Newen, 2018), particularly regarding sensorimotor coupling with sociocultural contexts, which are saturated with unobserved characteristics of interaction, that need to be accounted for within a mechanistic framework (De Jaegher and Di Paolo, 2012; Rojas-Líbano and Parada, 2020).

A promising avenue –within research on the relevance of both sensory-motor information (Wilson, 2002; Gallese, 2007; Shapiro, 2011) and the experiential context of their process of perception– has been opened by the corpus of theories of embodied, embedded, extended and enactive cognition (Varela et al., 2017), which challenged the assumption that nervous systems evolved for abstract thought (in terms of mere throughput processing) and rather did for the adaptive control of action (Semin and Smith, 2007); therefore, conceptual structure ought to be grounded in an experiential foundation specific to the sensory-motor system. However, as Eliasmith (2003) points out, research insight is still intertwined with metaphors specific for researchers' methodologies at the best, and I would argue that, at the worst, it usually remains at the level of object-oriented ontologies.

4E Cognition Explanations as Incommensurable Conceptual Metaphors

Philosophical bias, in the context of contemporary cognitive neuroscience, can be explained as the fundamental assumptions made between ontology (what is), epistemology (what can be known) and practical norms (how science should be practice regarding operative/operational concepts such as causality, probability and complexity, while following ideals like objectivity, reliability, validity, coherence, transparency and rationality). As Andersen et al. (2019) state, these "(. . .) count as biases because they skew the development of hypotheses, the design of experiments, the evaluation of evidence, and the interpretation of results in specific directions," although "Sometimes these assumptions are chosen deliberately and explicitly by the scientist, and used as auxiliary premises for theoretical purposes."

Furthermore, as Craver (2014) points out: "Not all of the facts in an ontic explanation are salient in a given explanatory context, and for the purposes of communication, it is often necessary to abstract, idealize, and fudge to represent and communicate which ontic structures cause, constitute, or otherwise are responsible for such phenomena."

It follows that mutually incommensurable (Kuhn, 1962; Feyerabend, 1970; Popper, 1996) conceptual metaphors (Lakoff and Johnson, 1980), although intrinsic to a phase of paradigmatic stability in scientific cycles of knowledge production, are also symptomatic of a lack of a completely coherent semiotic framework (Proni, 2015) that could account for currently unobserved characteristics of interaction, which saturate social contexts, remaining invisibilized because of these discursive practices: "(. . .) The linguistic entities that are called "explanations" are statements reporting the actual explanation. Explanations, in this (ontic) view, are fully objective and (. . .) no epistemically relativized (. . .)" (Salmon, 1989).

Models of cognition that have been informed by the 4E-Cognition (Wilson, 2002; Glenberg and Kaschak, 2003) epistemologies have failed regarding the conveyance of higher cognitive states and, even more so, social shared meaning and individual/group learning. Indeed, Zlatev (2007) has said regarding embodiment that "There are, however, three major unresolved issues within the current embodied turn in the sciences of the mind" and at least six within the language sciences (Ostarek and Huettig, 2019). The first was mentioned in passing already: there is not one but many different meanings behind the term embodiment, both between and within fields, and the corresponding theories are in general not compatible (Ziemke, 2003). There is no uniform concept of representation within "embodied cognition."

The central issue with the purely symbolic perspective has thus not been resolved through the embodied turn, as already put forward by Brette (2019): representing is not some kind of register or data structure that we use, but something we do, as "Items, memory, data, structure, etc. can do nothing relevant except influence process flow, and those influences can, in principle, always be built directly into the process organization" (Erdin and Bickhard, 2018).

IS A SCIENCE OF SOCIAL COGNITION CONCEIVABLE?

Perspectivalness

Social cognition demands the exploration of concepts like interiority and intersubjectivity, which have been held in distance from the possibility of being studied in an interactive way and primarily regarded as a mere contextual descriptors for individual mechanisms. Indeed, Frith (2008) has expressed that "mainly third-person aspects of social-cognitive processes have been considered" so far, even though, as Krakauer et al. (2017) have pointed out: "many have argued for the importance of second-person, participatory capabilities." They have gone as far as to claim that "Insofar as the goal of a neuroscience research question is to explain some behavior, be it a phenomenon form vision, communication, motor control, navigation, language, memory, or decision making, the behavioral research must be considered, for the most part, epistemologically *prior.*"

De Jaegher et al. (2010) had already argued that "the role of interactive and individual elements in social cognition must be systematically re-evaluated" although they concede, "that social cognition may occur in the absence of interaction."

To probe the access to others' intentions requires escaping an essentialist and universalizing model of theory of mind. Linguists hold that a child cannot proficiently learn to speak without this capacity (Robbins and Rumsey, 2008). Pauen (2012) suggests that knowing this "perspectivalness" directly enhances the ability to take the second-person perspective, which would essentially allow for epistemic replication to take place. Goldman's (2006) simulation theory had already posited as the central problem of imagining another mind's subjective experience the actual capacity for proper categorization of contextual information.

Agency

An analogous historical case to take into consideration is Dual Inheritance Theory (also known as gene-culture co-evolution), which effectively broadened the scope of what ought to be considered fit for description regarding the interplay of human physiology and cognition. Nonetheless, following the complexity in the notion of agency –as they put it: "Control is delegated to a system of poorly understood internal drives and rewards that direct the activity of the individual" (Cavalli-Sforza and Feldman, 1981)—DIT was limited to treating subjects simply as self-interested machines.

Nowadays, it is possible to account for several of the biases that undergo social learning and knowledge transmission, and thus attempt to quantify the chance over time of aspects of cognition within a mechanistic framework. This allows us to focus, for instance, on usage frequencies [i.e., regularizations (Reali and Griffiths, 2010)] of the "more richly structured" (Lieberman et al., 2007) aspects of language and to inquire if models of neutral selection can account for these behaviors. This could inform theories of cognition across all levels of information-processing systems (Marr, 1982; Pylyshyn, 1984), a prevailing need for which, according to Newberry et al. (2017) there is a consensus among several disciplines (i.e., neuroscience, artificial intelligence, linguistics, philosophy, psychology and anthropology).

Even if one were to continue tackling dimensions that involve representational content, there is an emerging claim within cognitive science of language that semantic composition is the primary structural selection factor over syntactic processing (Blank et al., 2016) and that there is a need for realistic models of what may have selected for their representations (Hauser et al., 2014).

POSITIVITY AND POSSIBILITIES OF REALISM

The epistemic question of how knowledge is being generated and how this is influencing the research results thus arises. Lende and Downey (2012) propose a holistic approach to improve onto this practice: by further strengthening the way we examine the relationship between recollection of objective data on changes in brain activity and the engagement of culture and individuals simultaneously. So called neuroanthropology places the brain at the center of discussions about human nature, following that "the nervous system is our most cultural organ." It emphasizes the interaction between the sociocultural milieu and its contingent sensory environment at the material level (i.e., in terms of brain percepts). Anthropology has long made the effort to posit the exploration of Self and Otherness within the scope of the cognitive sciences. Likewise, topics such as the representational requirements of cognition in their relation to the dynamic, circular and distributed causal structure of the brain have not been studied through second-person perspective or ethnographic methods yet, but have been limited to be described by the use of questionnaires at the most.

Within a world of causality, Mead (1962) concisely referred to affordances (Gibson, 1977) by commenting on their potentiality: "The chair invites us to sit down." Thus, they contribute to the emergence of meaning, since the response to the aforementioned invitation does not depend on cognitive representations alone but they come into play "through particular actions and projects of the subjective selves of the sentient entities" (Keane, 2013). These are central concepts of current robotics, artificial intelligence and information architecture upon which the ethnographic method has to shed some light; this possibility needs to be acknowledged for scientific advancement.

The conveyance of New Realism entails this legitimization. As we have learned specifically from linguistics and more broadly from 4E-Cognition that all structure is social in two ways: it exists through construction and acquires meaning through interpretation. Szwedek (2011) referred to "the ultimate source domain" –the physical– that needs to be cross-ontically mapped before any further higher cognitive metaphorization occurs, as "(...) before any entity can be assigned structure or orientation, it must be objectified first." It follows that we should learn that explorations on social cognition stemming from discursive practices are not entirely materially unobservable and are actually filled with relevant information –social cognitive affordances– which current working metaphors are not able to represent.

CONTINGENT COGNITIVE CONSTELLATIONS

We have described the issue of interaction dynamics having been approached to the point of conceptual saturation by cognitive neuroscience, by mainly making use of representation-oriented and inference-based perspectives. The point of overflow has been reached: current working metaphors within contemporary scientific discursive practices, even though informed through 4E Cognition, no longer address the contingency of sociocultural interaction. Nonetheless, embodiment and situationism themselves pointed toward information pervasively present throughout social contexts which still remains invisibilized and thus requires an ontic and epistemic turn to be accounted for.

Following De Lauretis (2004), who referred to theory as being invested in figuring out the now—i.e., the enigma of the world and argued in favor of theoretical inquiry by stating that "(...) thinking, however abstract, originates in an embodied subjectivity, at once over determined and permeable to contingent events," this invitation to a new turn toward realism is the attempt to gain specific insight into a contingent social cognition by way of observing the possible ways sensory constellations actually function and how these create thoroughly different but rich representations of the physical (Fluegge, 2003).

As there cannot be a single way of creating knowledge, only such a broad, socioculturally-founded yet materially based perspective will perhaps allow us to fill in the missing elements.

AUTHOR CONTRIBUTIONS

NAH conceptualized the present work and wrote the current version for publication.

FUNDING

The present work was supported by Comisión Nacional de Investigación Científica y Tecnológica (CONICYT, Chile), through National Doctoral Grant No. 21140102, as well as by FONDECYT Regular Grant No. 1150336. Funding sources (CONICYT, Chile) had no role in the production and development of this manuscript.

ACKNOWLEDGMENTS

The author would like to thank Prof. Dr. Bernardo Riffo, Dr. Mabel Urrutia and Prof. Mónica Véliz de Vos for their constant support throughout the emergence of this work. I would also like to thank the organizers of the profoundly inspiring conference Social cognition in humans and robots, held at the University Medical Center Hamburg-Eppendorf and organized by the FET Proactive H2020 project Socializing Sensorimotor Contingencies and the European Society for Cognitive Systems, without which this work had not come to be.

REFERENCES

Andersen, F., Anjum, R. L. and Rocca, E. (2019). Philsophical bias is the one bias that science cannot avoid. *Elife*:e44929. doi: 10.7554/eLife.44929

Barsalou, L. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660. doi: 10.1017/S0140525X99002149

Blank, I., Balewski, Z., Mahowald, K. and Fedorenko, E. (2016). Syntactic processing is distributed across the language system. *Neuroimage* 127, 307–323. doi: 10.1016/j.neuroimage.2015.11.069

Brette, R. (2019). Is coding a relevant metaphor for the brain? *Behav. Brain Sci.* 42:49. doi: 10.1017/S0140525X19001997

Cavalli-Sforza, L. L., and Feldman, M. (1981). *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton, NJ: Princeton University Press.

Craver, C. F. (2014). "The ontic account of scientific explanation," in *Explanation in the Special Sciences, Synthese Library (Studies in Epistemology, Logic, Methodology and Philosophy of Science)*, eds M., Kaiser, O., Scholz, D., Plenge, A. D. Hüttemann (Dordrecht: Springer), 367.

De Jaegher, H., and Di Paolo, E. (2012). The interactive brain hypothesis. *Front. Hum. Neurosci.* 6:163. doi: 10.3389/fnhum.2012.00163

De Jaegher, H., Di Paolo, E., and Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends Cogn Sci.* 14, 441–447. doi: 10.1016/j.tics.2010.06.009

De Lauretis, T. (2004). Statement due. *Crit. Inquiry* 30, 365–368. doi: 10.1086/421134

Eliasmith, C. (2003). *Neural Engineering: Unraveling the Complexities of Neural Systems*. Cambridge: MIT Press.

Erdin, H. O. and Bickhard, M. H. (2018). Representing is something that we do, not a structure that we "use": Reply to Gladziejewski. *New Ideas Psychol.* 49, 27–37. doi: 10.1016/j.newideapsych.2018.02.001

Feyerabend, P. K. (1970). *Problems of Empiricism*. UK: Cambridge University Press.

Fluegge, E. (2003). The consideration of personal sound space: toward a practical perspective on individualized auditory experience. *J. Sonic Stud.* (Leiden) 1:1–16. Available online at: http://journal.sonicstudies.org/vol01/nr01/a09

Frith, C. D. (2008). Social cognition. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 363, 2033–2039. doi: 10.1098/rstb.2008.0005

Gallese, V. (2007). Before and below theory of mind: Embodied simulation and the neural correlates of social cognition. *Philos. Trans. R. Soc. Biol. Sci.* 362, 659–669. doi: 10.1098/rstb.2006.2002

Gibson, J. J. (1977). "The theory of affordances," in *Perceiving, Acting and Knowing*, eds R. Shaw and J. Bransford (Hillsdale, NJ: Lawrence Erlbaum Associates).

Glenberg, A. M., and Kaschak, M. P. (2003). "The body's contribution to language," in *The Psychology of Learning and Motivation: Advances in Research and Theory*, Vol. 43, ed B. H. Ross (Elsevier Science), 93–126.

Goldman, A. I. (2006). *Simulating Minds: The Philosophy, Psychology and Neurosience of Mindreading*. Oxford: Oxford University Press.

Hauser, M. D., Yang, C., Berwick, R. C., Tattersall, I., Ryan, M. J., Watumull, J., et al. (2014). The mystery of language evolution. *Front. Psychol.* 5. doi: 10.3389/fpsyg.2014.00401

Keane, W. (2013). On spirit writing: materialities of language and the religious work of transduction. *J. R. Anthropol. Institute* 19, 1–17. doi: 10.1111/1467-9655.12000

Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. and Poeppel, D. (2017). Neuroscience needs behavior: correcting a reductionist bias. *Neuron* 93, 480–490. doi: 10.1016/j.neuron.2016.12.041

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press.

Lakoff, G. and Johnson, M (1980). *Metaphors We Live By*. Chicago, IL: University of Chicago Press.

Lende, D. H. and Downey, G. (2012). *The Encultured Brain: An Introduction to Neuroanthropology*. Cambridge: MIT Press. doi: 10.7551/mitpress/9219.001.0001

Lieberman, E., Michel J.-B., Jackson, J., Tang, T. and Nowak, M. A. (2007). Quantifying the evolutionary dynamics of language. *Nature* 449, 713–716. doi: 10.1038/nature06137

Marr, D. (1982). *Vision: A Computational Approach*. San Francisco, CA: W. H. Freeman and Company.

Mead, G. G. (1962). *Mind, Self, and Society: From the Standpoint of a Social Behaviorist*. Chicago, IL: University of Chicago Press.

Meteyard, L., Cuadrado, S. R., Bahrami, B. and Vigliocco, G. (2012). Coming of age: a review of embodiment and the neuroscience of semantics. *Cortex* 48, 788–804. doi: 10.1016/j.cortex.2010.11.002

Newberry, M. G., Ahern, C. A., Clark, R. and Plotkin, J. B. (2017). Detecting evolutionary forces in language chance. *Nature* 551, 223–226. doi: 10.1038/nature24455

Newen, A. (2018). The embodied self, the pattern theory of self, and the predictive mind. *Front. Psychol.* 9:2270. doi: 10.3389/fpsyg.2018.02270

Ostarek, M. and Huettig, F. (2019). Six challenges for embodiment research. *Curr. Direct. Psychol. Sci.* 28, 593–599. doi: 10.1177/0963721419866441

Pauen, M. (2012). The second-person perspective. *Inquiry* 55, 33–49. doi: 10.1080/0020174X.2012.643623

Popper, K. R. (1996). *The Myth of the Framework*. London: Routledge.

Proni, G. (2015). Umberto eco and charles peirce: a slow and respectful convergence. *Semiotica* 206, 13–35. doi: 10.1515/sem-2015-0021

Pylyshyn, Z. (1984). *Computation and Cognition*. Cambridge: MIT Press.

Reali, F., and Griffiths, T. L. (2010) Words as alleles: connecting language evolution with Bayesian learners to models of genetic drift. *Proc. R. Soc. B.* 277, 429–436. doi: 10.1098/rspb.2009.1513

Robbins, J. and Rumsey, A. (2008). Introduction: Cultural and linguistic anthropology and the opacity of other minds. *Anthropol. Q.* 81, 407–420. doi: 10.1353/anq.0.0005

Rojas-Líbano, D. and Parada, F. J. (2020). Body-world coupling, sensorimotor mechanisms and the ontogeny of social cognition. *Front. Psychol*. 10:3005. doi: 10.3389/fpsyg.2019.03005

Salmon, W. C. (1989). "Four decades of scientific explanation," in *Scientific Explanation. Minnesota Studies in the Philsophy of Science,* eds P. Kitcher and W. C. Salmon (Minnesotta: University of Minnesota Press), 183–219.

Semin, G. R. and Smith, E. R. (2007). Situated social cognition. *Curr. Direct. Psychol. Sci.* 16, 132–135. doi: 10.1111/j.1467-8721.2007.00490.x

Shapiro, L. (2011). Embodied Cognition. New York: Routledge Press.

Storch, M. and Tschacher, W. (2014). *Embodied Communication: Kommunikation beginnt im Körper im Kopf*.Bern: Huber.

Szwedek, A. (2011). The ultimate source domain. *Rev. Cogn. Ling.* 9, 341–366. doi: 10.1075/rcl.9.2.01szw

Urrutia, M. and de Vega, M. (2012). Lenguaje y acción: Una revisión actual a las teorías corpóreas. RLA. *Rev. Lingüíst. Teórica Apl.* 50, 39–67. doi: 10.4067/S0718-48832012000100003

Varela, F. J., Thompson, E. and Rosch, E. (2017). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge: MIT Press. doi: 10.7551/mitpress/9780262529365.001.0001

Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev.* 9, 625–636. doi: 10.3758/BF03196322

Ziemke, T. (2003). "What's that thing called embodiment?" *Proceedings of the 25th Annual meeting of the Cognitive Science Society, Vol.* 6.

Zlatev, J. (2007). "Embodiment, language and mimesis," in *Body, Language, Mind*, *Vol. 1*, eds T. Ziemke, J. Zlatev, and R. Franck, (Berlin: De Gruyter Morton) 297–337.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 06 October 2020
doi: 10.3389/fnhum.2020.559793

Collaboration Variability in Autism Spectrum Disorder

Maria Blancas[1,2,3], Giovanni Maffei[1,2], Martí Sánchez-Fibla[3], Vasiliki Vouloutsi[1,2] and Paul F. M. J. Verschure[1,2,4]**

[1] Synthetic Perceptive Emotive Cognitive Systems (SPECS) Laboratory, Institute for Bioengineering of Catalonia (IBEC), Barcelona, Spain, [2] Barcelona Institute of Science and Technology (BIST), Barcelona, Spain, [3] Department of Information and Communication Technology, Pompeu Fabra University, Barcelona, Spain, [4] Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain

This paper addresses how impairments in prediction in young adults with autism spectrum disorder (ASD) relate to their behavior during collaboration. To assess it, we developed a task where participants play in collaboration with a synthetic agent to maximize their score. The agent's behavior changes during the different phases of the game, requiring participants to model the agent's sensorimotor contingencies to play collaboratively. Our results (*n* = 30, 15 per group) show differences between autistic and neurotypical individuals in their behavioral adaptation to the other partner. Contrarily, there are no differences in the self-reports of that collaboration.

Keywords: autism, prediction, collaboration, sensorimotor contingencies, neurodiversity

Edited by:
Leonhard Schilbach,
Ludwig Maximilian University of Munich, Germany

Reviewed by:
Murat Perit Çakir,
Middle East Technical University, Turkey
Lihui Wang,
Shanghai Jiao Tong University, China

****Correspondence:***
Maria Blancas
maria.blancas@upf.edu
Paul F. M. J. Verschure
paul.verschure@specs-lab.com

Specialty section:
This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience

Received: *07 May 2020*
Accepted: *03 September 2020*
Published: *06 October 2020*

Citation:
Blancas M, Maffei G, Sánchez-Fibla M, Vouloutsi V and Verschure PFMJ (2020) Collaboration Variability in Autism Spectrum Disorder.
Front. Hum. Neurosci. 14:559793.
doi: 10.3389/fnhum.2020.559793

INTRODUCTION

Autism spectrum disorder (ASD) is a neurodevelopmental disorder whose main impact falls in two domains: persistent deficits in social communication and restricted, repetitive patterns of behavior or interest (*DSM-5* American Psychiatric Association, 2013). ASD has been linked to a deficit in prediction abilities and to the fact that feedback is more prominent compared to feed-forward anticipatory control (Schmitz et al., 2003; Sinha et al., 2014; Wang et al., 2014).

Recent research (Sinha et al., 2014) suggests that a prediction deficit present since early development (Prediction Impairment in Autism, PIA hypothesis, in Sinha et al., 2014) could cause the diversity of expression of the autism syndrome. This theory divides the prediction difficulties among insistence on sameness, sensory hypersensitivities, interacting with dynamic objects, theory of mind, and islands of proficiency. Insistence on sameness represents repetitive actions and thoughts, inflexible adherence to routines, resistance to change, and ritualized patterns of verbal or non-verbal behavior. Sensory hypersensitivities refer to the sensory abnormalities (like hypersensitivity to bright light) experienced by individuals in the spectrum, however, these abnormalities are not caused by abnormally enhanced sensation. Individuals in the autism spectrum also have difficulties with theory of mind (that is, inferring mental states to others and ascribing causes to observations about a person through the connection of previous with current behavior), which can cause deficit-adjusting behavior to suit different social situations. Finally, they can exhibit enhanced abilities in strongly rule-based domains (known as islands of proficiency). These domains, like mathematics, musical performance, or calendar calculations, are strongly rule-based, which minimizes uncertainty.

Individuals with ASD show attenuated top-down prior expectations, which leads them to rely more on bottom-up sensory signals. They thus experience hypersensitivity, enhanced perception and sensation, and sensory overload (Mitchell and Ropar, 2004). Consequently, this dependence on perceptual evidence merged with impairments in contextualizing sensory evidence impedes understanding actions, and predicting social intentions. Nevertheless, individuals with ASD do not show difficulties in perceiving social stimuli, but rather in using them to update internal models of social interaction, what leads to impairments in social abilities (South et al., 2012; D'Cruz et al., 2013; Robic et al., 2015).

The so-called social symptoms encompass deficits in social interaction and communication. These poor "social-specific" priors compromise their interaction with others, as ASD individuals have difficulties in coping with the uncertainty that comes with social behaviors (Chambon et al., 2017). Acting together with another partner requires considering and integrating both one's own and the partner's next action. This planning of cooperative actions, although less studied, is also considered an aspect of sensorimotor control (Sebanz et al., 2006).

Sensorimotor integration can be defined as the brain process allowing response to specific demands of the environment by executing voluntary motor behavior (Machado et al., 2010). Planning and executing a simple movement require sensory feedback, to effectively coordinate movement while acting. Thus, sensorimotor approaches consider perception and action as a united process. This interaction between action and perception must be highlighted in sensorimotor approaches, as they are not seen as separate processes. On the contrary, actions are conferred an integral function for perception to explain cognitive functions.

To consider an anticipatory effect as reflecting prospective sensorimotor control, an action has to differ depending on the subsequent one (Rosenbaum et al., 2013; Ansuini et al., 2015). Sensorimotor contingencies (SMCs) can be seen technically as forward models that predict the expected sensory changes given a certain set of movements. Knowledge of SMCs allows an agent to simulate potential outcomes of behavioral alternatives. Impairments in sensorimotor integration could lead to ineffective use of sensory feedback in, for example, movement correction. As a result, the individual could face difficulties in coordination and sensory reactivity.

The main brain areas associated with sensorimotor integration are the cerebellum (Paulin, 1993; Glickstein, 1998) and the basal ganglia (Nagy et al., 2006; Chukoskie et al., 2013). It is not surprising, therefore, the significant differences found in these specific areas of autistic patients. For example, previous research showed a lower number of Purkinje cells in the cerebellum (Bauman and Kemper, 2005; Amaral et al., 2008) and a decreased volume in the basal ganglia (Estes et al., 2011) in ASD individuals as compared to typically developed ones.

The cerebellum is suggested to control the anticipatory and predictive adjustments of motor programs (Koziol et al., 2012). Its pathways link sensory signals to motor areas in the brain (Glickstein, 1998), which have a pivotal role in controlling and coordinating movement (Paulin, 1993). Research on autism has provided ample evidence that the cerebellum is among the most frequently disrupted brain regions in ASD (Palmen et al., 2004; Courchesne et al., 2005), with persistent differences in volume emerging since the first 2 years of life (Hashimoto et al., 1995; Stanfield et al., 2008). Studies suggest that ASD is characterized by alterations of the brain's inference on the causes of socially relevant signals, and this lack of ability to predict actions of other individuals stems from cerebellar dysfunctions (Schmitz et al., 2003; Sinha et al., 2014; Wang et al., 2014).

The basal ganglia play a functional role in sensory integration and motor control (Nagy et al., 2006). This area, reciprocally connected to the cerebellum (Chukoskie et al., 2013), has previously been claimed to be different in individuals with autism. For example, it has a lower volume than typical brains (Estes et al., 2011), and one of its areas, the striatum, shows larger functional connectivity in individuals with autism (Di Martino et al., 2011). Previous research has shown weak connectivity between sensory and motor brain areas in individuals with autism (Oldehinkel et al., 2019). These findings are consistent with the sensory symptoms (such as hypersensitivity) experienced in ASD. They are also in line with work showing out of sync interactions between visual and motor regions in individuals in the spectrum.

The aforementioned alterations in sensory input and motor execution could play a pivotal role in autism. The available evidence seems to suggest that autism shows widespread disturbances in sensorimotor behavior (Haswell et al., 2009; Rinehart and Mcginley, 2010; Cook et al., 2013; Gowen and Hamilton, 2013; Thompson et al., 2017). Along similar lines, self-reports about sensorimotor behavior coming from people in the spectrum provide further evidence on sensory alteration and over-responsivity (Kern et al., 2006; Ben-Sasson et al., 2009; Tavassoli et al., 2014).

Some examples of sensorimotor alterations in ASD comprise impaired motor processing and higher detection of unattended changes compared to neurotypical individuals. There is support presenting these impairments in movement and sensory responsivity not as a peripheral feature of autism, but as a fundamental cause of the social and communicative impairments seen in the condition (Leary and Hill, 1996; Hilton et al., 2007; Reynolds et al., 2011; Matsushima and Kato, 2013). Sensorimotor difficulties in autism are associated with the development and maintenance of social impairments characteristic of the disorder. Integrating sensory information from the environment is required to plan and execute movement effectively, to, altogether, carry on proper social reciprocity.

The relation between sensorimotor impairments and social deficits in autism suggests impairments in the coupling of perceptual and social cues. More specifically, ASD individuals may encounter difficulties using the sensorimotor contingencies exhibited by another agent to predict the agent's behavior. Thus, this work focuses on the evaluation of the coupling of perceptual and social cues based on sensorimotor interaction and the ability to predict another agent's behavior. More specifically, we aim to assess how predictive abilities affect collaborative interaction and how they differ between ASD and Typically Developed (TD) individuals.

To do so, we devised a predictive game task where participants collaborate with a synthetic agent that displays different behavioral patterns expressed through sensorimotor contingencies. The proposed task is an adaptation of the game of Pong, where players in collaboration with a synthetic agent need to intercept a falling target (see the following section for more information). To succeed in this task, players need to identify and learn the social characteristics of the agent. Doing so will allow them to use this information during the interaction, and to later adapt more efficiently when the task becomes uncertain. As the agent's behavior is based on sensorimotor information, we hypothesize that ASD individuals will show deficits in successful social predictive/anticipatory skills. To assess the differences in prediction between ASD and TD players, we look at aspects of adaptive collaborative skills by analyzing the interaction of the players with the AI agent of the game, how the interaction evolves during the task, and how it relates to the participants' understanding of the other agent's characteristics. More specifically, we study partner monitoring and how it affects the covered space and look at the mutual influence between the player and the AI-controlled agent.

We hypothesize that:

- Participants in the autism spectrum will show slower and less adaptation to the other agent than neurotypical ones.
- Participants in the autism spectrum will show less adaptation to the other agent when the task becomes more uncertain.
- Participants in the autism spectrum will show more variable behavior than neurotypicals.

THE SCENARIO

The Task

The purpose of the current study is to evaluate how goal-oriented coordination between partners could be achieved through sensorimotor adaptation. To do so, we designed a collaborative multiplayer version of the game of Pong: a computer version of a 2D tennis where two players try to intercept falling targets from the top before they hit the ground by moving their paddles at the bottom of the screen. The paddles move on the same horizontal line and can push each other but cannot switch sides. In this game, one player is AI-controlled, and the other is a human. **Figure 1** represents an example of the proposed scenario.

For this task, we considered a collaborative team task like playing tennis doubles, where each player should cover a maximal part of their field so that all targets return to the opponent's side. Targets sometimes fall in the middle part of the field, thus in a zone where both players could intercept the target. The location of the target was randomly selected from a uniform distribution of possible angles, and the pace of the target drop was uniform across all trials. The velocity of the artificial player was controlled and the same across all trials and the velocity of the participant depended on their motion on the trackpad. A player can be characterized by the area they cover and intervene, given the target's direction. Typically, in a game of two, the area covered by each player is half of the playable area. However, more active players may sometimes overpass their area to try and catch ambiguous targets directed toward the middle area. Collisions with the other agent were penalized by subtracting a point, and participants were informed about the penalty before beginning the task. To evaluate whether the synthetic agent's behavior and predictability can influence the humans' behavior, we varied the playing styles of the agent.

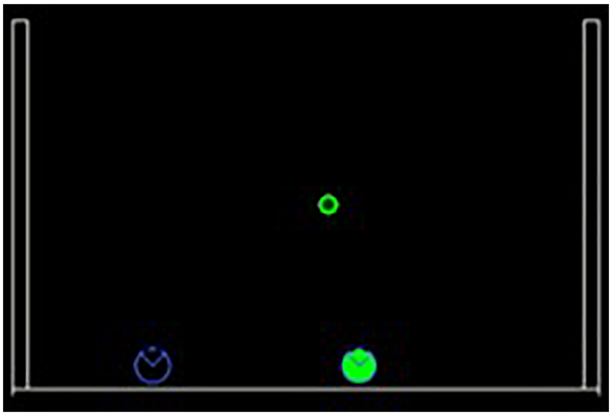


FIGURE 1 | Example of a trial during the task. Targets fall from the top of the screen and players need to intercept them before they hit the ground. The player on the left (blue) is controlled by a human and the one on the right (green) is the synthetic agent. This example represents interaction with the "Middle" agent.

The AI-controlled player differs in the way it approaches the target and the area in which it will intervene, resulting in three different agents: "Wider," "Narrower," and "Middle." A "Middle" agent will try to intercept any targets that fall within its half of the space and has a 0.5 probability of intercepting an ambiguous target that falls in the middle. A "Wider" agent will try to intercept the target and overpass its area to try and catch a target even if the target's position is not ambiguous. In contrast, a "Narrower" agent would try to intercept the target without overpassing its area; in fact, it would cover a space that is smaller than half of the overall space. The next section explains in more detail how the agent's behavior is obtained.

The Point of Social Subjective Equality

To measure the collaboration between the human player and the AI player, we introduce the Point of Social Subjective Equality (PSSE). The PSSE can be computed for every two players and all possible target trajectories. This measurement is an analytical measure of collaboration (i.e., social affordance gradient) that defines the probability of going for the target depending on the target's position (**Figure 2**, left). Therefore, the PSSE is the point where each player has the same probability of going to intercept the target (**Figure 2**, right) and is an extension of the Point of Subjective Equality (PSE) (Stoloff et al., 2011) to a socially collaborative task. PSE represents the point where there is an equal probability of using any of the two hands to reach a target (presented from left to right circularly in front of the participant). Thus, the Point of Social Subjective Equality indicates how a player is relying or not on the partner, invading or not the partner's area of the field while intercepting targets in the horizontal range. In short, it is the point where a player has an equal probability of intercepting the target or letting their

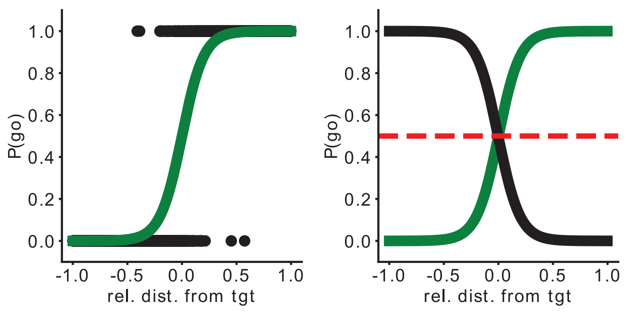


FIGURE 2 | The left image represents the probability to go for the target depending on the player's relative distance to the target. The right image depicts the Point of Social Subjective Equality (PSSE). Here, the green line represents the AI agent (in this case, the "Middle" one) and the black line, a simulated perfectly matched participant. The red dashed line represents the moment when both agents have the same probability of going for the target. The x-axis represents the relative distance of the agent from the target; the y-axis, the probability it has to go for the target.

partner intercept it. To calculate it, we first calculate the relative distance from the player to the target (Eq. 1), that is, the difference between one player to the target and the other player to the target. After that, we fit a sigmoid function with the distance to the target (*rel_dist*), a constant factor (*k*), and a bias value (*b*) representing the behavior of each of the agents (Eq. 2). We estimated the parameters of the PSSE (*k*, *bias,* and *rel_dist*) by running a logistic regression using sklearn[1]. To our knowledge, this is the first time such a direct behavioral measure of collaboration is introduced.

$$rel\_dist = \frac{(|p_t - p_{p2}| - |p_t - p_{p1}|)}{w} \tag{1}$$

Representation of the relative distance (rel_dist). p_t represents the position of the target, p_{p2} represents the position of the other agent, p_{p1} represents the position of the participant, and w represents the width of the (game) screen.

$$PSSE = \frac{1}{1 + e^{-(k \times rel\_dist + b)}} \tag{2}$$

Representation of the Point of Social Subjective Equality (PSSE). b represents each partner's bias, *rel_dist* is the relative distance from the target, and k is a constant factor ($k = 20$).

Based on the PSSE, two complementary partners would intercept the target with the same probability ($P = 0.5$, **Figure 2**, right), whereas any shift would indicate a lack of balance between the partners. As mentioned previously, participants play with three different AI agents, and we modulated their behavior based on this shift of the interception point. Our three proposed agents, namely "Middle" (M), "Wider" (W), and "Narrower" (N), have therefore different probabilities of intercepting the target. More specifically, the "Middle" agent has a 0.5 probability of going for an ambiguous target (when the target falls in the center of the arena). A "Wider" agent is more prone to invade the space of the participant; therefore, the curve of the probability to intercept the target based on the target's location would fall toward the left part of the space. In contrast, the "Narrower" agent is more prone to stay in its half of the space and allow the participant to enter the AI agent's space to catch the target. Consequently, the curve would fall toward the right part of the space. Thus, if we split the playable area into two equally sized sides, one for the participant and the other for the synthetic agent, a "Middle" agent would cover only its 50% of the space, while the "Wider" would cover more than 50% and the "Narrower," less. **Figure 3** provides an example of the representation of the curve for each AI agent. The agents were programmed to catch the target following a pre-defined strategy (M, W, or N). Consequently, if a participant decided to leave the target to the artificial agent, the agent's behavior would depend on the predefined strategy and therefore the position of the target and the relative positions of the two players. Thus, there would be cases where the ball would be intercepted by the artificial agent and others where it would be missed, however, the PSSE sigmoid function would not be affected by the movement of the human player. The coefficient and the intercept of this curve will allow us to assess participants' adaptation to the other agent.

MATERIALS AND METHODS

Participants

The ASD participants recruited for the study had previously been diagnosed as autistic, meeting the *DSM-5* criteria for level 1 of autism ("Requiring support;" American Psychiatric Association, 2013) ($N = 15$, one female, age: 18.67 ± 2.4). This criterion comprises difficulties in initiating social interactions and switching between activities. This group was recruited in the Educa Friends center[2], an educational support service part of the Friends Foundation, focused on providing support to

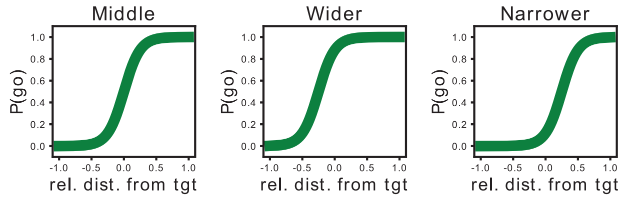


FIGURE 3 | Representation of the curve of the probability to intercept the target based on the target's location for each of the three proposed agents. From left to right: the curve of the "Middle" agent lies in the middle as both the synthetic agent and the human player have the same probability of intercepting the target. In contrast, the curve of the "Wider" synthetic agent is slightly skewed toward the left, as this agent will enter the space of the human participant. In contrast, the "Narrower" agent's curve is skewed toward the right; this agent has a higher probability of staying toward its half of the space and allowing the human participant to intercept the target.

---

[1] https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

[2] https://fundaciofriends.org/es/servicios/educafriends/

high-level functioning ASD individuals. All participants had a normal or corrected-to-normal vision and were not color blind. Participants were matched for handedness (all of them were right-handed) and almost matched for age (same mean, different standard deviation) and gender (only one female more in the TD group). The typically developed participants were recruited in a high school of Barcelona and the campus of the Polytechnic University of Catalunya, and their age matched those with ASD ($N = 15$, two females, age: 18.38 ± 1.06). Written informed consent was obtained for all participants (for the ones under the age of 18, parental written informed consent was obtained too). The study was approved by the local ethical committee (Parc de Salut del Mar).

Apparatus

Participants sat at a viewing distance of (approximately) 50 cm from a 27-inch monitor that operated at a resolution of 1,920 × 1,080. The monitor was part of an All-in-One desktop computer connected to a touchpad and a keyboard. The task was generated using Python and the PyGame library, and participants controlled their avatar using the touchpad. There was no auditory feedback during the task. **Figure 4** depicts the setup used.

Experimental Procedure

All participants were provided with an information sheet that contained the explanation of the task and a consent form they had to sign before beginning the experiment. For the underaged participants, information sheets and consent forms were given to both participants and their parents/legal tutors. Before the main task, participants filled in a small questionnaire with demographics and the frequency of playing video games. As mentioned earlier, the task is a Pong adaptation, and the goal is to intercept falling targets. The task was performed in a computer using a touchpad and consisted of three main phases. In "Phase 1," participants played alone for one block. In all phases, each block consisted of 150 trials. The number of trials was decided after running a pilot study with other participants (excluded from this sample), both in the spectrum and neurotypical. To our understanding, 150 trials are enough to cover the probability distribution for each participant, as players do not really cover all the horizontal range of the screen. According to the design of the task, both players go for the target within their field and only need to decide whether to go or not when the target falls in the middle. Thus, the focus is on the center of the distribution, narrowing the range of interest by requiring fewer samples to build a probability distribution for each participant. In "Phase 2," participants played for one block with each of the three AI players (in total three blocks), and finally, in "Phase 3," participants played for one block with all agents. The order of the three AI players in "Phase 2" was randomized and for each type of the three AIs. "Phase 3" was used to assess the social predictive abilities of the participants, as they had to interact with a random agent in every trial (counterbalanced so there were 50 trials with each agent). Each of the three agents was depicted in a different color (Neutral, blue; Wider, green; and Shy, White). Color choices were made arbitrarily. Players' positions were initialized to the center of their side at the beginning of each trial. Participants were instructed to avoid hitting the other agent and were penalized with one point less if they did.

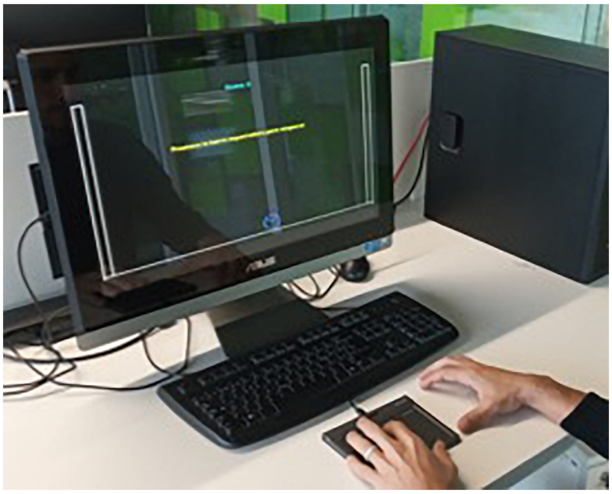


FIGURE 4 | Representation of the setup used in the task. Participants sit in front of a computer screen where the game was displayed. Participants controlled the motion of their avatar using a touchpad.

In this study, we used behavioral data, questionnaires, and interviews as instruments to collect information about the participants' behavior and perception of the task. Between each of the blocks, participants had to answer questions in a tablet. The questions involved perceived collaboration and predictability of the target and the other agent and engagement. To answer, participants had to rate each of them on a Likert scale from 1 to 5. At the end of the task, we carried out a semi-structured interview to assess the perceived differences between agents, followed by a debriefing session. **Figure 5** represents the experimental protocol. In total, the whole experiment took around 30 min and was conducted in Spanish or Catalan, depending on the preferred language of the participant.

Data Collection

To evaluate the behavioral and perceptual differences between the two groups, we collected data gathered from the logs of the game (behavioral), questionnaires, and short interviews (perceptual). More specifically, from the logs of the game, we obtained in a trial-by-trial basis the *performance* (one point if either the participant or artificial player intercepted the target), the *identifier* of the artificial player (M, W, N), and *the position of the player, target, and agent*. The last three positions (the positions of both agents during the last three time frames in the trial) allowed us to obtain the *PSSE* measure (by analyzing their relative distance when one of them intercepted the target), as explained in Eq. 1.

The between blocks questionnaire allowed us to assess participants' perception of the task and the artificial agent. In all blocks, participants evaluated *task engagement* and *target predictability*. In "Phase 2" and "Phase 3," where the artificial

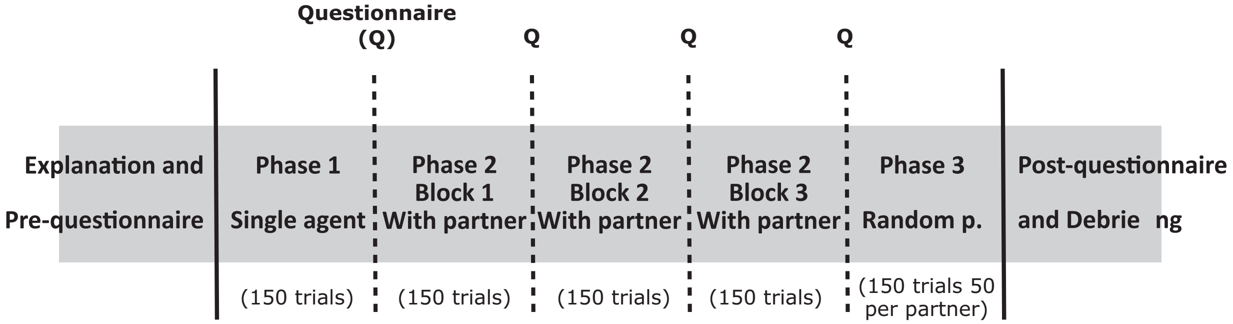


FIGURE 5 | Representation of the experimental protocol. First, participants are introduced to the task and fill in a short questionnaire. The task comprises three phases, which contain one, three, and one blocks (of 150 trials per block), respectively. In "Phase 1," participants play alone. In "Phase 2," participants randomly play with each of the AI agents for one block (three blocks in total). In "Phase 3," participants are presented with a random agent in each trial (50 trials per agent). Self-reports on perceived collaboration, engagement, and agents' and target predictability are presented between phases/blocks. Finally, participants fill in a short questionnaire and undergo an interview and debriefing.

player was introduced, participants also evaluated the agent's *predictability* and *collaboration*. All items were reported on a 5-point Likert scale. At the end of the experiment, participants were also asked to report if they thought the other player was a human or a computer, adding "I do not know" as a possible answer.

Finally, the short-structured interview at the end of the experiment allowed us to assess with further detail participants' perception of the task and the other agents. More specifically, we asked participants to report on the overall perceived difficulty of the task and report on how they perceived the other player. Here, participants could describe the other player and if they have identified any differences between the blocks. Furthermore, we asked participants to assess the difficulty of "Phase 3" of the task (in each trial, participants played with a random AI) and report if they followed any strategy.

RESULTS

The following results have been analyzed in Python, using the following libraries: NumPy, JSON, math, scipy, and sklearn. In order to choose the statistical tests used in this analysis, we ran normality tests in the variables. The intercepts of the PSSEs during "Phase 2" had a normal distribution, so parametric tests were used (One-Way ANOVA, in this case). The mean squared errors in "Phase 2" did not show a normal distribution, so non-parametric tests were used (Mann-Whitney U).

Behavioral Results

We defined performance as the ratio of caught targets out of the 150 of each phase. A Mann Whitney U test ($U = 892448.0$, $p = 0.433$) showed that there were no significant differences between the two groups (ASD: median: 1.0, MAD: 0; TD: median: 1.0, MAD: 0) in *performance* during "Phase 1" (when participants played alone). Thus, possible differences in "Phase 2" and "Phase 3" should not be related to their performance when playing alone.

Participants in the Autism Spectrum Showed More Variable Behavior Than Neurotypicals

To assess participants' adaptation to the artificial player, we calculated the Point of Social Subjective Equality (PSSE). First, we look at the two groups' behavior in "Phase 2," where we take into account all trials with each agent per block. To analyze the differences between groups and agents, we calculated the differences between the coefficients among groups for the same agent, and among agents for the same group. There were no differences between agents in their coefficients in none of the groups (**Figure 6**). In terms of intercepts, there were significant differences between agents in both the ASD [One-way ANOVA (6,749) = 5.68, $p = 0.007$] and TD group [One-way ANOVA (6,749) = 10.83, $p < 0.001$]. More specifically, an independent samples t-test showed differences in the ASD group were between the Middle and Narrow agents [$t(4{,}499) = -2.11$, $p = 0.04$] and the Narrow and Wider agents [$t(4{,}499) = 3.46$, $p = 0.002$]; and in the TD between the Middle and Narrow agents [$t(4{,}499) = -5.17$, $p < 0.001$] and the Narrow and Wider agents [$t(4{,}499) = 4.01$, $p < 0.001$].

The lack of significant difference between slopes could mean that, generally, both groups adapted in a similar way. Nevertheless, as we can see in **Figure 6**, participants in the ASD group showed a higher probability of going (∼0.25) with a relative distance larger than 0. This means that they had more tendency to go toward the target than the TD group (which probability at that time was around 0.07), even when they should not. The differences in intercept represent the adaptation of each group to the specific agent they were playing with. In the next sections, we will quantify the variability of each group and their behavioral changes with respect to the other agent.

To assess the variability among participants in each group, we calculated the mean squared error between each participant and the general mean. To do so, we first calculated the general mean of the coefficients extracted from the data points obtained in all trials in all blocks from "Phase 2" in both groups. From that, we calculated the average of those data points and obtained a

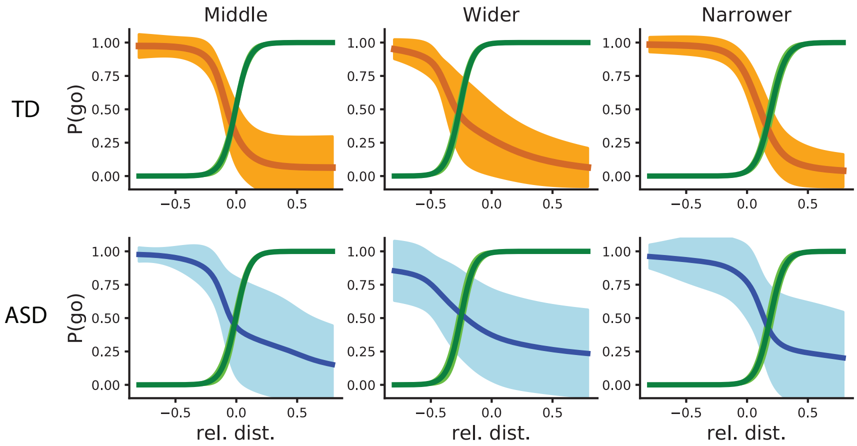


FIGURE 6 | PSSE intersections from the TD and ASD groups (top and orange and bottom and blue, respectively). From left to right: the PSSE curve of the AI agent (depicted in green) and that of the participants, depicted in orange (TD) and blue (ASD), when playing with a "Middle" (N), "Wider" (A), and "Narrower" (S) agent respectively. The bright colors represent standard deviation, while the darker and thicker line represents the mean.

representative mean squared error per group (ASD: 0.08 ± 0.08; TD: 0.001 ± 0.001). A Mann Whitney U test was used to analyze the differences between groups against the general mean ($U = 9.0$, $p < 0.001$). Moreover, when assessing the variability inside of each group (that is, the variability compared to the mean of their group), the difference is even higher ($U = 0$, $p < 0.001$). The U equal to zero signifies that all the mean squared errors in the ASD group are greater compared to all the ones in the TD group.

Participants in the Autism Spectrum Showed Slower Adaptation to the Artificial Agent Than Neurotypical Ones

To further understand the two groups' adaptation, we then looked at a possible evolution in time of the PSSE, and more specifically, whether early (50 first trials) and late (50 last trials) trials differed between the groups in "Phase 2." To do so, we analyzed the shift in PSSE for each of the agents. We used the "Middle" agent as a baseline and subtracted from it the shift for the "Wider" and "Narrower" agents. Like this, we could calculate how much the participants' behavior changed when encountering the "Wider" and "Narrower" agents. As we can observe in **Figure 7**, we found statistically significant differences between the two groups for the "Wider" agent in the early trials ($U = 65.0$, $p = 0.042$) but not the late trials. We did not find any statistically significant differences between groups for the "Narrower" agent in both early and late trials.

Participants in the Autism Spectrum Showed Less Adaptation to the Other Agent When the Task Became More Uncertain

During "Phase 2" we have shown that healthy subjects acquired an *ad hoc* behavioral strategy (i.e., PSSE shift) from the interaction with each individual agent and that the adaptation process was more pronounced in healthy subjects compared to control. During "Phase 3" we aim at assessing whether this strategy can be correctly retrieved when the subjects interact with each agent in a randomized order. We hypothesized that the ASD group will be less able to retrieve a correct strategy, potentially due to the reduced ability to form an internal model of the partner. To do so, for every subject we compute the PSSE associated to each agent during "Phase 2."

In "Phase 2," participants played for one block with each of the three AI agents. In contrast, in "Phase 3," participants encountered a random AI player in each trial for one block. As mentioned earlier, the characteristic that distinguishes the agents' behavior is the color, and if players have not made the color association with the agent's behavior, "Phase 3" becomes more uncertain. Here, we wanted to assess how much the players' behavior in "Phase 3" matches that of "Phase 2" when playing with the same agent during each of the blocks. To do so, we ran a logistic regression using participants' behavior during "Phase 2" as our "training data," and compared against their behavior during Phase 3, which was used as "testing data." PSSE for each agent is described by a logistic function with constant k and intercept i. We further group the trials from "Phase 3" according to the agent type and extract, similarly to "Phase 2," the probability of the subject to go for the target or to let the partner go ($p = 1$ and $p = 0$, respectively). Finally, we compute for every agent how accurately the parameters of the PSSEs from "Phase 2" describe the behavior (i.e., probability of going for the target) observed in "Phase 3." The rationale is that high accuracy of the model from "Phase 2" in describing the behavior of "Phase 3" would confirm the hypothesis that a behavioral strategy tight to each individual agent has been learned and can be correctly retrieved.

In the left panel of **Figure 8**, we show the mean accuracy matrix for the control group and the ASD group. This is obtained by computing for every subject the accuracy of each PSSE agent model (predicted) in describing the data of each agent during

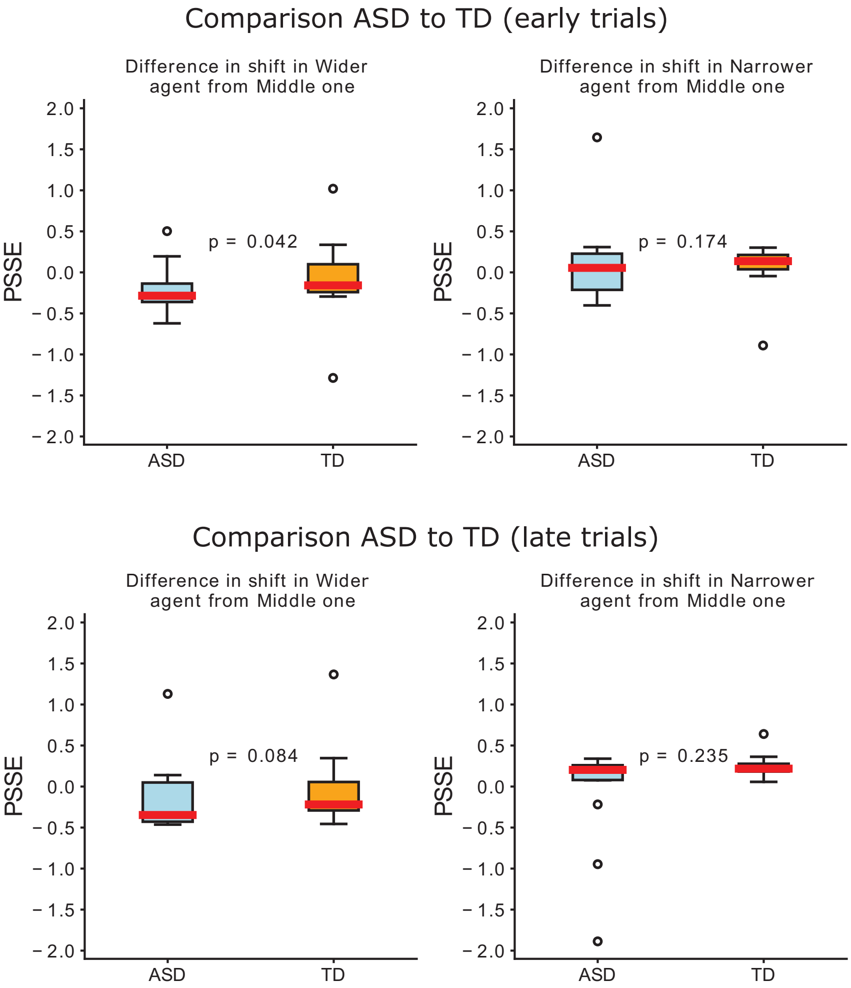


FIGURE 7 | Differences between groups in the difference in shift per agent (compared to the "Middle" one). The upper plots represent early (0–50) trials and the lower plots, late (100–150) trials. Blue represents the ASD group; orange, the TD; and the red line represents the median.

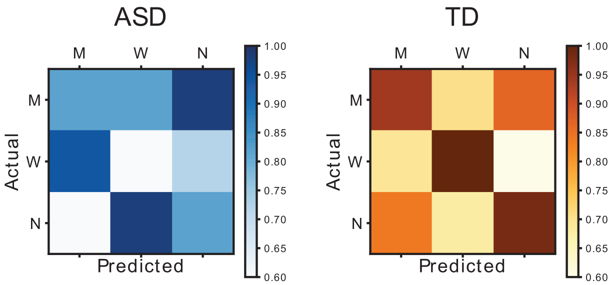


FIGURE 8 | Matrix representing the relationship between Phase 2 (actual) and Phase 3 (predicted) behavior. The blue matrix represents the group in the autistic spectrum, while the orange one represents data from the neurotypical group.

"Phase 3." This generates a set of 3 × 3 matrices that are further averaged for each group. This result suggests that the participants in the neurotypical group (right, orange) behaved in the same way in both phases, with a mean accuracy score of 0.97). However, the participants in the ASD group (left, blue) did worse in properly matching their behavior to the one in the previous phase (mean accuracy score of 0.74). A Mann Whitney-*U* test showed significant differences between the accuracy for both groups in matching Phases 2 and 3 behavior ($U = 0$, $p = 0.03$). These results could suggest that participants in the TD group developed a model of the other player during "Phase 2" that they later used to adapt their behavior in "Phase 3"; participants in the ASD group failed to do so.

No Differences in Perception of the Task Between Groups, Only by Perceiving the Other Agent as Human or Synthetic

As previously mentioned, participants had to answer a short questionnaire between blocks. More specifically, participants evaluated target predictability, engagement (in all blocks), as well as agent predictability and collaboration (in the blocks where the

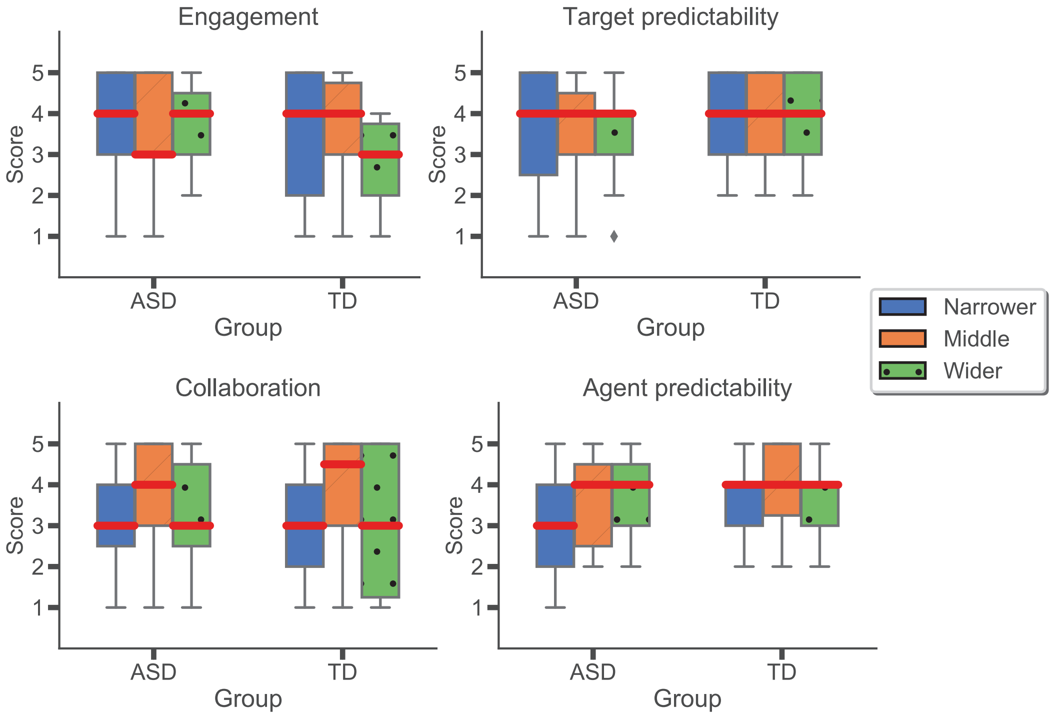


FIGURE 9 | Differences between agents and groups in perceived agent predictability, target predictability, engagement, and collaboration. Blue represents the results for the "Narrower" agent; orange, for the "Middle" agent; and green, from the "Wider" one. Red lines represent means.

AI agent was introduced). There were no significant differences in engagement or target predictability between "Phase 1" and the rest of the task. In "Phase 2," participants rated each agent at the end of each respective block (**Figure 9**). Results suggest no statistically significant differences between groups in any of the dimensions. Nonetheless, participants in the ASD group seemed to feel more engaged with the task than the neurotypical group. Additionally, we observe higher variability in the ASD group when evaluating target predictability. In contrast, the TD group evaluated the target's predictability similarly in all three blocks. Regarding collaboration, both groups reported the "Middle" agent as the most collaborative one. Finally, we could observe differences in the perceived agent predictability, where the "Narrower" agent was perceived as less predictable by the ASD group than the TD one.

Finally, at the end of "Phase 3," participants reported if they were interacting with a human or a computer. There were no significant differences between the ASD and TD groups as to how many participants thought they were playing with a human or a computer. Interestingly, if we divide participants into two new groups (those that thought the other agent was a human and those who thought it was a computer), we observe differences in perceived collaboration (**Figure 10**). Participants that thought the other agent was a human perceived it as significantly more collaborative (Human: median 4.0, MAD: 0.0; Synthetic: median: 3.0, MAD: 1.0; Mann-Whitney U: 37.0, $p = 0.01$). More specifically, when comparing among agents (by assessing the answers during Phase 2, where participants provided self-reports for each of the agents separately), the agent that was perceived as more collaborative was the "Middle" (Human: median 5.0, MAD: 0; Computer: median: 3.0, MAD: 2.0; Mann-Whitney U: 33.0, $p = 0.009$), followed by the "Wider" (Human: median 4.0, MAD: 1.0; Computer: median: 3.0, MAD: 1.0; Mann-Whitney U: 43.0, $p = 0.04$) and the "Narrower" (Human: median 4.0, MAD: 1.0; Computer: median: 3.0, MAD: 2.0; Mann-Whitney U: 43.5, $p = 0.04$).

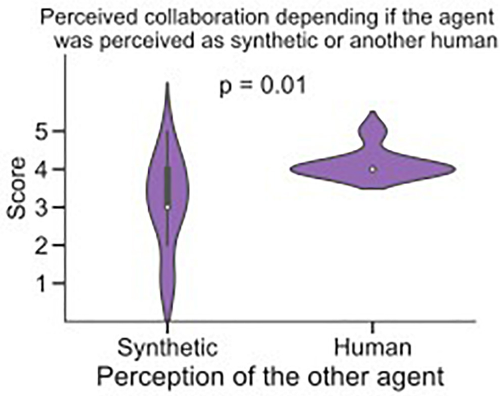


FIGURE 10 | Differences in perceived collaboration between the participants that perceived the AI agent as synthetic or as a human player. The white dot represents the median.

Participants in the ASD Group Focused on Movement to Describe the Other Agent

After the last questionnaire in "Phase 3," participants underwent a short structured interview, which lasted ∼10 min. First, participants were asked to report the difficulty of the task and how well they performed. We later asked them to comment and describe the other agent they interacted with. In the case where participants reported differences between the agents, we also asked them to provide a short description for each agent. When participants described the agent(s) they interacted with, we identified the following common features: movement (how fast/slow the agent was perceived), field use (how much of the field the agent was using), color (the color of the agent), and collaboration (how collaborative the agent was perceived). **Figure 11** depicts the frequency of use of these characteristics to differentiate between the agents (sometimes, more than one per subject). In the ASD group, the most commented characteristic was the agents' movements (53%), followed by their field use (33%), perceived collaboration (13%), and color (6%). In the TD group, the frequency of use of the characteristics is more homogeneous. Here the most frequent characteristics are collaboration and field use (38%) followed by color and movement (30%). Moreover, one participant in the TD group differentiated between the agents by their perceived performance.

The two groups mainly differed in the type of characteristics they used to describe the other agent. Participants in the ASD used more personality-related terms to describe the behavior of the other agent ("it's Narrower," "it's more selfish") than neurotypical participants, who used a more performance-related vocabulary ("it was playing well," "it was taking my targets"). When asked about "Phase 3," in which they played with a random AI agent in each trial, participants in the ASD group communicated an added difficulty caused by larger uncertainty. Some reported not knowing if the other agent would go or not for the target; others reported that the task required more focus ("You never know what can happen or how will the other player react. You had to be more focused"). Only one subject reported a relationship between the agent's color and its behavior and using it to decide to go or not for the target. In contrast, in the TD group, more participants reported using color to identify the agent and act accordingly. For both groups, "Phase 3" was perceived as more complicated and confusing than the others.

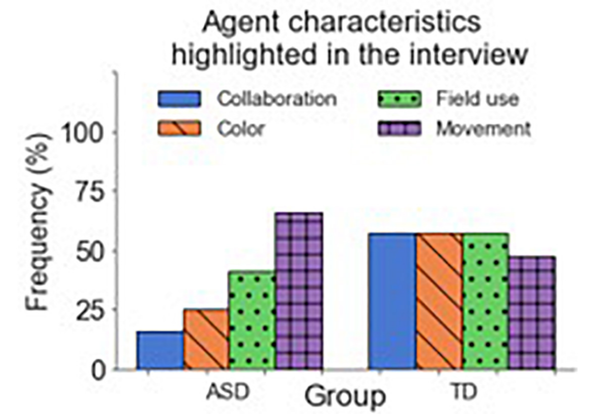


FIGURE 11 | Frequency of characteristics commented about during the interviews. Blue represents collaboration-related characteristics; purple, color-related ones; green represents characteristics related to field use; and orange, characteristics related to movement. The sum of the frequencies inside each group surpasses 100% because some subjects highlighted more than one characteristic.

DISCUSSION AND CONCLUSION

The main purpose of this study was to evaluate the ability to predict another agent's behavior based on sensorimotor interaction and how these predictive abilities affect collaborative interaction and differ between ASD and Typically Developed (TD) individuals. We created a task where participants had to learn the behavioral characteristics (as exhibited by sensorimotor contingencies) of a synthetic agent and collaborate with the agent to maximize reward. Each player controlled an avatar, and the goal was to intercept falling targets. To assess collaboration, we developed the Point of Social Subjective Equality (PSSE) that calculates the probability of a player of going for the target given the target's position. Finally, we examined possible perceptual differences regarding the task between the two groups.

Discussion on Differences in Behavior

As we observe larger individual differences between participants in the autism spectrum (compared to neurotypicals), we hypothesized that participants in the autism spectrum would show more variable behavior than neurotypicals during the task. Our analysis of the differences in variability between the ASD and TD groups suggests that, indeed, the ASD group showed larger variability compared to the TD individuals.

Social impairments associated with sensorimotor difficulties are a characteristic of the disorder, and we assumed that ASD individuals would encounter difficulties in predicting the AI player's behavior. Thus, we hypothesized that participants in the autism spectrum would show slower and less adaptation to the other agent than neurotypical ones. To assess this, we analyzed the differences between groups in adaptation to the other agent during Phase 2, and we showed differences in adaptation between groups in early trials but not in late ones, showing differences in adaptation timing. Our results show differences in the behavior of neurotypical and ASD individuals when playing with the three different synthetic agents. We observe the ability to converge to a complementary PSSE in the case of the control group. However, we do not observe the same with the ASD participants. Furthermore, we assessed the online adaptation to the artificial player by looking at the differences in errors between early and late trials among groups. Our results seem to reflect a more accurate adaptation in the neurotypical group than in the ASD.

Finally, as ASD individuals seem to find difficulties when a task is uncertain, we postulated that they would fail to predict the behavior of the AI agent correctly and, therefore, adapt less to the AI agent compared to the typically developed group. By comparing the participants' behavior with each agent in "Phase 2" and "Phase 3," we can assess whether they applied previously acquired information from the sensorimotor contingencies of the AI agent ("Phase 2") to a more uncertain task ("Phase 3") and predict the agent's behavior. Our results suggest differences in the prediction of the agent's behavior. More specifically, TD individuals were able to develop a better model of the artificial player in "Phase 2" and apply that information to adapt their behavior in "Phase 3," while participants in the ASD failed to do so.

These results suggest that neurotypical individuals can adapt their behavior according to the AI player and converge to an optimal game strategy by observing the sensorimotor patterns of their partner. In contrast, ASD patients seem to lack this ability, suggesting an impairment of socSMCs, possibly due to their lower predictive skills (Schmitz et al., 2003; Sinha et al., 2014; Wang et al., 2014).

Discussion on Differences in Perception

To understand possible perceptual differences between the two groups, we looked at the questionnaires provided to the participants after the completion of each block, and the short structured interview at the end of the task. Participants at the end of each block reported how they evaluated the task, the target, and the other player in terms of engagement, predictability, and collaboration. Although we did not find any statistical significance in any of the items, participants in the ASD group seemed to perceive the task as overall more engaging than the TD group. Participants in the TD group perceived the task as less engaging when interacting with the "Wider" agent. When evaluating the agent's predictability, the "Narrower" agent was perceived as less predictable by the ASD group compared to the TD. Playing with the different agents did not seem to affect target predictability in the TD group. However, we observe higher variability in the reported target predictability in the ASD group when playing with the "Narrower" and "Middle" agents. In terms of collaboration, both groups rated the "Middle" agent as more collaborative than the "Narrower" and "Wider." Despite a lack of significance, our results provide possible insights on perceptual differences regarding the tasks' characteristics with respect to the agent's behavior. However, more data needs to be collected.

At the end of the task, we asked participants to report whether they thought they interacted with another human or a computer. We found no significant differences between the two groups. The agent was perceived as significantly more collaborative by participants that thought they were playing with a human instead of a computer. More specifically, the "Middle" agent was rated 50% more collaborative when participants thought it was another human. Indeed, according to Turing's test (Turing, 1950), the behavior of a machine can be confused with that of a human.

Finally, the short-structured interview allowed us to assess further the perceived differences of the agents between the ASD and TD groups. The main differences arise from the characteristics used to describe the agents. Participants in the ASD group mainly commented on the agents' movements, followed by their field use. The agents' color was the characteristic less commented about. In contrast, the TD group differentiated between the agents by almost equally exploiting all three characteristics. The focus on movement as the main differentiating characteristic is something to be expected from the ASD group, as individuals in the spectrum tend to focus on moving objects. Moreover, the fact that almost no subject in the ASD group commented on the agents' color as a significant trait could support the idea of the lack of model generation. If the agent's color was a characteristic that could help participants predict its behavior, it would be unnecessary to consider it if no model was being created.

GENERAL DISCUSSION

The contributions of this study are two-fold. On the one hand, we formulated and introduced the Point of Social Subjective Equality (PSSE), a concept that allowed us to model the behavior of both humans and artificial agents in a collaborative task. By observing the PSSE, we quantified the degree of behavioral adaptation and how it can be modulated based on the variation of sensorimotor contingencies of the synthetic agent. On the other hand, this study demonstrated how collaborative behavior could implicitly emerge and be modulated through the observation of sensorimotor patterns of the partner.

Our behavioral analysis showed lower and slower adaptation to the artificial player by the ASD group. Similar results were found in Lieder et al. (2019), where participants in the autistic spectrum showed lower and slower adaptability in the task than their neurotypical counterparts. However, previous studies examining sensorimotor planning in individuals with ASD have yielded conflicting results. Some studies indicate an impairment in prospective control in ASD (Hughes, 1996; Scharoun and Bryden, 2016). In contrast, other studies showed no significant differences (Hamilton et al., 2007; van Swieten et al., 2010).

The larger variability in behavioral results of the ASD group is also present in the self-reported data. Nevertheless, the perceived predictability and collaboration during the task showed no differences between groups in these measures. Interestingly, the differences in the behavioral data but not in the self-reports raise the question of self-awareness. Could that be due to a lack of metacognition or due to a coping mechanism? Unfortunately, our current data do not allow us to answer this question, and further studies need to be conducted.

CONCLUSION AND FURTHER STEPS

In conclusion, this study adds to the literature possible ways of measuring collaboration through sensorimotor contingencies, and how this collaboration is impaired in individuals with ASD. While this study provides a preliminary insight, several limitations need to be discussed. First, further studies with larger sample sizes are needed to better control for individual differences. Furthermore, it is important to note that our study lacks female participants, as the main general users of the ASD center we collaborated with were males. This is in line with the larger occurrence of autism in male individuals compared to female ones. Despite these limitations, this study proposes a simple (and non-invasive) method to evaluate the predictive abilities of individuals in the autism spectrum. To do so, more data would be needed, as the main limitation of this study is the weakness of its statistical power.

Possible uses of this application would go in the line of an environment where the user could train their social abilities in a controlled and adaptive way. The system could be used to improve the abilities of non-neurotypical people by training their predictive skills. Like this, individuals in the spectrum could not only train their tracking of moving objects and predict their trajectories but also train their reading and understanding of non-verbal cues. The task offers the possibility of merging these two types of prediction (related to objects and social interaction), in a game-like manner.

To the moment, the PSSE has not been contrasted with any kind of diagnostic tool for ASD. In the future, a validation of the PSSE measurement in comparison with a screening tool could allow for a stronger claim on distinguishing between these two groups. However, at this point, we do not claim that it can be either a diagnostic tool or a tool to be able to distinguish between the two groups, but we highlight the possibility.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Parc de Salut Del Mar. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

MB ran the experiments and data analysis, and wrote the manuscript. GM ran the PSSE code and data analysis. MS-F ran the PSSE idea and code. GM, MS-F, VV, and PV reviewed the manuscript. All authors read and approved the manuscript.

FUNDING

This work received funding from the European Research Council under the European Union's Seventh Framework Programme socSMCs (Grant No. EC, H2020-641321). MS-F was supported by Spanish National Grant INSOCO-DPI2016-80116-P.

ACKNOWLEDGMENTS

We would like to thank the participants, their parents, and the therapists and teachers from the centers for their participation in this research.

REFERENCES

Amaral, D. G., Schumann, C. M., and Nordahl, C. W. (2008). Neuroanatomy of autism. *Trends Neurosci.* 31, 137–145. doi: 10.1016/j.tins.2007.12.005

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders*, 5th Edn. doi: 10.1176/appi.books.9780890425596

Ansuini, C., Cavallo, A., Bertone, C., and Becchio, C. (2015). Intentions in the brain: the unveiling of Mister Hyde. *Neuroscientist* 21, 126–135. doi: 10.1177/1073858414533827

Bauman, M. L., and Kemper, T. L. (2005). *The Neurobiology of Autism*. Baltimore, MD: Johns Hopkins University Press.

Ben-Sasson, A., Hen, L., Fluss, R., Cermak, S. A., Engel-Yeger, B., and Gal, E. (2009). A meta-analysis of sensory modulation symptoms in individuals with autism spectrum disorders. *J. Autism Dev. Disord.* 39, 1–11. doi: 10.1007/s10803-008-0593-3

Chambon, V., Farrer, C., Pacherie, E., Jacquet, P. O., Leboyer, M., and Zalla, T. (2017). Reduced sensitivity to social priors during action prediction in adults with autism spectrum disorders. *Cognition* 160, 17–26. doi: 10.1016/j.cognition.2016.12.005

Chukoskie, L., Townsend, J., and Westerfield, M. (2013). *Motor Skill in Autism Spectrum Disorders: A subcortical View. International Review of Neurobiology*, 1st Edn, Vol. 113. Amsterdam: Elsevier.

Cook, J. L., Blakemore, S. J., and Press, C. (2013). Atypical basic movement kinematics in autism spectrum conditions. *Brain* 136, 2816–2824. doi: 10.1093/brain/awt208

Courchesne, E., Redcay, E., Morgan, J. T., and Kennedy, D. P. (2005). Autism at the beginning: microstructural and growth abnormalities underlying the cognitive and behavioral phenotype of autism. *Dev. Psychopathol.* 17, 577–597. doi: 10.1017/S0954579405050285

D'Cruz, A.-M., Ragozzino, M. E., Mosconi, M. W., Shrestha, S., Cook, E. H., and Sweeney, J. A. (2013). Reduced behavioral flexibility in autism spectrum disorders. *Neuropsychology* 27, 152–160. doi: 10.1037/a0031721

Di Martino, A., Kelly, C., Grzadzinski, R., Zuo, X. N., Mennes, M., Mairena, M. A., et al. (2011). Aberrant striatal functional connectivity in children with autism. *Biol. Psychiatry* 69, 847–856. doi: 10.1016/j.biopsych.2010.10.029

Estes, A., Shaw, D. W. W., Sparks, B. F., Friedman, S., Giedd, J. N., Dawson, G., et al. (2011). Basal ganglia morphometry and repetitive behavior in young children with autism spectrum disorder. *Autism Res.* 4, 212–220. doi: 10.1002/aur.193

Glickstein, M. (1998). *Cerebellum and the Sensory Guidance of Movement*. Chichester: John Wiley & Sons.

Gowen, E., and Hamilton, A. (2013). Motor abilities in autism: a review using a computational context. *J. Autism Dev. Disord.* 43, 323–344. doi: 10.1007/s10803-012-1574-0

Hamilton, A. F. D. C., Brindley, R. M., and Frith, U. (2007). Imitation and action understanding in autistic spectrum disorders: how valid is the hypothesis of a deficit in the mirror neuron system? *Neuropsychologia* 45, 1859–1868. doi: 10.1016/j.neuropsychologia.2006.11.022

Hashimoto, T., Tayama, M., Murakawa, K., Yoshimoto, T., Miyazaki, M., Harada, M., et al. (1995). Development of the brainstem and cerebellum in autistic patients. *J. Autism Dev. Disord.* 25, 1–18. doi: 10.1007/BF02178163

Haswell, C. C., Izawa, J., Dowell, R., Mostofsky, S., and Shadmehr, R. (2009). Representation of internal models of action in the autistic brain. *Nat. Neurosci.* 12, 970–972. doi: 10.1038/nn.2356

Hilton, C., Graver, K., and LaVesser, P. (2007). Relationship between social competence and sensory processing in children with high functioning autism spectrum disorders. *Res. Autism Spectr. Disord.* 1, 164–173. doi: 10.1016/j.rasd.2006.10.002

Hughes, C. (1996). Brief report: planning problems in autism at the level of motor control. *J. Autism Dev. Disord.* 26, 99–107. doi: 10.1007/BF02276237

Kern, J. K., Trivedi, M. H., Garver, C. R., Grannemann, B. D., Andrews, A. A., Savla, J. S., et al. (2006). The pattern of sensory processing abnormalities in autism. *Autism* 10, 480–494. doi: 10.1177/1362361306066564

Koziol, L. F., Budding, D. E., and Chidekel, D. (2012). From movement to thought: executive function, embodied cognition, and the cerebellum. *Cerebellum* 11, 505–525. doi: 10.1007/s12311-011-0321-y

Leary, M. R., and Hill, D. A. (1996). Moving on: autism and movement disturbance. *Ment. Retardat.* 34, 39–53.

Lieder, I., Adam, V., Frenkel, O., Jaffe-Dax, S., Sahani, M., and Ahissar, M. (2019). Perceptual bias reveals slow-updating in autism and fast-forgetting in dyslexia. *Nat. Neurosci.* 22, 256–264. doi: 10.1038/s41593-018-0308-9

Machado, S., Cunha, M., Velasques, B., Minc, D., Teixeira, S., Domingues, C. A., et al. (2010). Sensorimotor integration: basic concepts, abnormalities related to movement disorders and sensorimotor training-induced cortical reorganization. *Rev. Neurol.* 51, 427–436.

Matsushima, K., and Kato, T. (2013). Social interaction and atypical sensory processing in children with autism spectrum disorders. *Hong Kong J. Occup. Ther.* 23, 89–96. doi: 10.1016/j.hkjot.2013.11.003

Mitchell, P., and Ropar, D. (2004). Visuo-spatial abilities in autism: a review *Infant Child Dev.* 13, 185–198. doi: 10.1002/icd.348

Nagy, A., Eördegh, G., Paróczy, Z., Márkus, Z., and Benedek, G. (2006). Multisensory integration in the basal ganglia. *Eur. J. Neurosci.* 24, 917–924. doi: 10.1111/j.1460-9568.2006.04942.x

Oldehinkel, M., Mennes, M., Marquand, A., Charman, T., Tillmann, J., Ecker, C., et al. (2019). Altered connectivity between cerebellum, visual, and sensory-motor networks in autism spectrum disorder: results from the EU-AIMS longitudinal european autism project. *Biol. Psychiatry Cogn. Neurosci. Neuroimag.* 4, 260–270. doi: 10.1016/j.bpsc.2018.11.010

Palmen, S. J. M. C., Van Engeland, H., Hof, P. R., and Schmitz, C. (2004). Neuropathological findings in autism. *Brain* 127, 2572–2583. doi: 10.1093/brain/awh287

Paulin, M. G. (1993). The role of the cerebellum in motor control and perception. *Brain Behav. Evol.* 41, 39–50. doi: 10.1159/000113822

Reynolds, D., Sammons, P., De Fraine, B., Townsend, T., and Van Damme, J. (2011). Educational effectiveness research (EER): a state of the art review. *Sch. Effect. Sch. Improv.* 43, 197–230. doi: 10.1080/09243453.2014.885450

Rinehart, N., and Mcginley, J. (2010). Is motor dysfunction core to autism spectrum disorder?. *Dev. Med. Child Neurol.* 52:697. doi: 10.1111/j.1469-8749.2010.03631.x

Robic, S., Sonié, S., Fonlupt, P., Henaff, M. A., Touil, N., Coricelli, G., et al. (2015). Decision-making in a changing world: a study in autism spectrum disorders. *J. Autism Dev. Disord.* 45, 1603–1613. doi: 10.1007/s10803-014-2311-7

Rosenbaum, D. A., Chapman, K. M., Coelho, C. J., Gong, L., and Studenka, B. E. (2013). Choosing actions. *Front. Psychol.* 4:273. doi: 10.3389/fpsyg.2013.00273

Scharoun, S. M., and Bryden, P. J. (2016). Anticipatory planning in children with autism spectrum disorder: an assessment of independent and joint action tasks. *Front. Integr. Neurosci.* 10:29. doi: 10.3389/fnint.2016.00029

Schmitz, C., Martineau, J., Barthélémy, C., and Assaiante, C. (2003). Motor control and children with autism: deficit of anticipatory function? *Neurosci. Lett.* 348, 17–20. doi: 10.1016/S0304-3940(03)00644-X

Sebanz, N., Bekkering, H., and Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends Cogn. Sci.* 10, 70–76. doi: 10.1016/j.tics.2005.12.009

Sinha, P., Kjelgaard, M. M., Gandhi, T. K., Tsourides, K., Cardinaux, A. L., Pantazis, D., et al. (2014). Autism as a disorder of prediction. *Proc. Natl. Acad. Sci. U.S.A.* 111, 15220–15225. doi: 10.1073/pnas.1416797111

South, M., Newton, T., and Chamberlain, P. D. (2012). Delayed reversal learning and association with repetitive behavior in autism spectrum disorders. *Autism Res.* 5, 398–406. doi: 10.1002/aur.1255

Stanfield, A. C., McIntosh, A. M., Spencer, M. D., Philip, R., Gaur, S., and Lawrie, S. M. (2008). Towards a neuroanatomy of autism: a systematic review and meta-analysis of structural magnetic resonance imaging studies. *Eur. Psychiatry* 23, 289–299. doi: 10.1016/j.eurpsy.2007.05.006

Stoloff, R. H., Taylor, J. A., Xu, J., Ridderikhoff, A., and Ivry, R. B. (2011). Effect of reinforcement history on hand choice in an unconstrained reaching task. *Front. Neurosci.* 5:41. doi: 10.3389/fnins.2011.00041

Tavassoli, T., Hoekstra, R. A., and Baron-Cohen, S. (2014). The sensory perception quotient (SPQ): development and validation of a new sensory questionnaire for adults with and without autism. *Mol. Autism* 5:29. doi: 10.1186/2040-2392-5-29

Thompson, A., Murphy, D., Dell'Acqua, F., Ecker, C., McAlonan, G., Howells, H., et al. (2017). Impaired communication between the motor and somatosensory homunculus is associated with poor manual dexterity in autism spectrum disorder. *Biol. Psychiatry* 81, 211–219. doi: 10.1016/j.biopsych.2016.06.020

Turing, A. M. (1950). I.—computing machinery and intelligence. *Mind* 59, 433–460. doi: 10.1093/mind/LIX.236.433

van Swieten, L. M., van Bergen, E., Williams, J. H. G., Wilson, A. D., Plumb, M. S., Kent, S. W., et al. (2010). A test of motor (not executive) planning in developmental coordination disorder and autism. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 493–499. doi: 10.1037/a0017177

Wang, S. S. H., Kloth, A. D., and Badura, A. (2014). The cerebellum, sensitive periods, and autism. *Neuron* 83, 518–532. doi: 10.1016/j.neuron.2014.07.016

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 12 November 2020
doi: 10.3389/frobt.2020.503452

Time-Aware Multi-Agent Symbiosis

Michail Maniadakis, Emmanouil Hourdakis, Markos Sigalas, Stylianos Piperakis, Maria Koskinopoulou and Panos Trahanias*

Institute of Computer Science, Foundation for Research and Technology – Hellas (FORTH), Heraklion, Greece

Contemporary research in human-machine symbiosis has mainly concentrated on enhancing relevant sensory, perceptual, and motor capacities, assuming short-term and nearly momentary interaction sessions. Still, human-machine confluence encompasses an inherent temporal dimension that is typically overlooked. The present work shifts the focus on the temporal and long-lasting aspects of symbiotic human-robot interaction (sHRI). We explore the integration of three time-aware modules, each one focusing on a diverse part of the sHRI timeline. Specifically, the Episodic Memory considers past experiences, the Generative Time Models estimate the progress of ongoing activities, and the Daisy Planner devices plans for the timely accomplishment of goals. The integrated system is employed to coordinate the activities of a multi-agent team. Accordingly, the proposed system (i) predicts human preferences based on past experience, (ii) estimates performance profile and task completion time, by monitoring human activity, and (iii) dynamically adapts multi-agent activity plans to changes in expectation and Human-Robot Interaction (HRI) performance. The system is deployed and extensively assessed in real-world and simulated environments. The obtained results suggest that building upon the unfolding and the temporal properties of team tasks can significantly enhance the fluency of sHRI.

Keywords: human robot interaction (HRI), artificial time perception, eterogeneous multi-agent planning, autonomous systems, collaborative task execution

Edited by:
Daniel Polani,
University of Hertfordshire,
United Kingdom

Reviewed by:
Annika Lübbert,
University Medical Center
Hamburg-Eppendorf, Germany
Alessandro Umbrico,
Italian National Research Council, Italy

****Correspondence:***
Michail Maniadakis
mmaniada@ics.forth.gr

Specialty section:
This article was submitted to
Human-Robot Interaction,
a section of the journal
Frontiers in Robotics and AI

Received: *08 October 2019*
Accepted: *29 September 2020*
Published: *12 November 2020*

Citation:
Maniadakis M, Hourdakis E, Sigalas M, Piperakis S, Koskinopoulou M and Trahanias P (2020) Time-Aware Multi-Agent Symbiosis. Front. Robot. AI 7:503452. doi: 10.3389/frobt.2020.503452

1. INTRODUCTION

Fluent, symbiotic Human-Robot Interaction (sHRI) is an important, yet challenging problem in robotics research as evidenced by the increasing number of published works (Rosenthal et al., 2010; Fernando et al., 2014; Liu et al., 2016; Riccio et al., 2016) and review papers (Coradeschi and Loutfi, 2008; Green et al., 2008; Carrillo and Topp, 2016; Tsarouchi et al., 2016). Despite the significant resources devoted in sHRI, the majority of existing systems consider mainly the spatial aspects of the world without encapsulating the concept of the time dimension. As a result, contemporary research has largely concentrated on enhancing robotic sensory, perceptual, and motor capacities, assuming short-term and nearly momentary interaction between agents (Das et al., 2015; Baraglia et al., 2016; Devin and Alami, 2016; Churamani et al., 2017). Still, human-machine confluence encompasses inherent temporal aspects that are often considered only implicitly in robotic applications, with clear negative effects regarding the integration of artificial agents into human environments. In example, robotic agents face difficulties in distinguishing between the entities involved in different past events or implement reasoning on past event sequencing, cannot feel rush or adapt to human temporal expectations and cannot effectively plan not only how, but also when tasks should be accomplished (Wilcox et al., 2012). Our recent work has addressed artificial temporal cognition, with a focus on human-like time representations and duration processing mechanisms for robots (Maniadakis et al., 2009, 2011; Maniadakis and Trahanias, 2012, 2015).

Interestingly, besides the fact that several cognitive architectures have considered for robotic systems over the last years (Langley et al., 2009; Rajan and Saffiotti, 2017; Kotseruba and Tsotsos, 2018), the notion of time is often represented rather implicitly in the knowledge base, without a clear view on the past, the present and the future of the robot life. For example, environment state changes are typically stored in a flat atemporal domain, being unable to distinguish between yesterday and a month before. The present work introduces a new cognitive framework that clearly separates between the well known notions of "past," "present," and "future," which are widely adopted by humans in their daily activities.

More specifically, three important phases of human-robot interaction can be easily identified in which time has a major role. These regard (i) the representation and memorization of past experiences on a temporally rich domain to facilitate time-informed reasoning in forthcoming sessions, (ii) the perception of the temporal features of evolving real-world procedures to support action coordination with other agents in the environment, and (iii) the planning of actions to facilitate timely accomplishment of goals given the temporal constraints and the dynamic unfolding of multi-agent collaboration. Targeting the implementation of time-aware robotic cognitive systems, we have developed computational modules addressing complementary cognitive skills along the past, present and future disciplines mentioned above (Maniadakis et al., 2016b, 2017; Sigalas et al., 2017a,b; Hourdakis and Trahanias, 2018).

In this paper we present the implementation of a composite sHRI system, that comprises the aforementioned time-aware cognitive modules. The composite system (a) exploits past experiences to reason about current human needs, (b) monitors and analyzes the ongoing human activity to infer the completion time of human tasks and the user's performance profile on the task, and (c) plans synergistic robot activities properly adapted to the human profile and the progress of the task in order to accurately satisfy human expectations. The paper summarizes the integration of the time-aware cognitive modules emphasizing mostly on their interactions and the beneficial features they bring to the composite system.

To demonstrate the validity of the proposed approach, the composite system is deployed in the real world and is assessed in a complex multi-agent interaction scenario that involves two robots and a human. A series of experiments with real humans showed that complementary to the embodiment of cognitive systems (i.e., link robot actions to body characteristics), the "entiment" of robotic cognition to the temporal context of sHRI (i.e., take into account when things happened or should happen) facilitates the coordination of robot behavior with the dynamic unfolding of the sHRI scenario. Overall, the use of time-informed robotic cognition facilitates the seamless integration of artificial agents in the real world, enhancing their ability to respond more accurately, flexibly, and robustly, in full accordance to the human expectations and needs.

The rest of the paper is organized as follows. Section 2 outlines related wok on the subject of time cognition, including issues of memory, temporal predictions and time-informed planning. Section 3 outlines the proposed interaction scenario that is employed in the current work, while section 4 provides details on the implementation of the composite system and the individual components. Section 5 presents a detailed experimental evaluation of the system in a set of human-centered experiments, while section 6 concludes the paper and discusses further research directions on the subject.

2. LITERATURE REVIEW

Over the years, a number of cognitive robotic architectures have considered the implementation of high-level cognitive functions taking into account temporal information, such as the constraints on the timing of tasks (Alami et al., 1998). For example, the deliberation for the completion of multiple and diverse robot tasks can be implemented based on six types of robot functions, namely planning, acting, observing, monitoring, goal-reasoning, and learning (Ingrand and Ghallab, 2017), where the need for timing robot activities has been also considered.

The broader field of human robot interaction has been significantly facilitated by the integration of modules which provide robust solutions on well-studied problems in the field of robotics. For example, Lemaignan et al. (2017) proposed a practical implementation for social human-robot interaction combining geometric reasoning, situation assessment, knowledge acquisition, and representation of multiple agents, for human-aware task planning. Churamani et al. (2017) built a human-robot interaction module to engage personalized conversations in order to teach robots to recognize different objects. Devin and Alami (2016) developed a framework which allows robots to estimate other agents' mental states e.g., goals, plans and actions and take them into account when executing human-robot shared plans. Das et al. (2015) proposed another framework for human-robot interaction based on the level of visual focus of attention. The latter was implemented on a Robovie-R3 robotic platform in order to interact with visitors in a museum. Adam et al. (2016) implemented a framework for physical, emotional and verbal human-robot interaction on a NAO robot.

Nevertheless, in these works the temporal dimension of human-robot interaction has not been adequately considered, since the focus of the relevant implementations was on the spatial aspects of task completion. As a result, the implemented systems are unable to develop a wider conception of the timeline linking the past, the present, and the future. The present work contributes to fill this gap by proposing an integrated system that directly considers the temporal characteristics of sHRI in order to realize long-term, timely and fluent cooperation of humans and robots. Previous works related to the components of the composite system are reviewed below.

2.1. Knowledge Representation and Reasoning

Robotic systems that naturally interact with humans for long periods should be equipped with the ability to efficiently store and manage past memories, as well as with the ability to exploit past experiences to predict future outcomes. Still, the temporal aspects of a robotic memory system have not yet been adequately examined, with most systems using flat, non-timed memories to assimilate past experiences. Accordingly, events that occurred at different past moments can be hardly distinguished, which results in poor performance in sHRI scenarios.

A common issue when encoding past events regards the management of the stored information, given the constantly increasing storage space over time. Memory forgetting (or decay), is a biologically inspired memory mechanism (Hardt et al., 2013) which may cope with this issue. The Decay Theory (Altmann and Gray, 2000) dictates that information stored in memory tends to "fade out" and/or eventually be forgotten as time passes. Prior attempts to computationally implement memory forgetting (Ho et al., 2009; Biswas and Murray, 2015) fail to dynamically adapt to variations in task requirements and, thus, are not suitable to support long-term sHRI. A dynamic memory system is proposed by the Time-Based Resource-Sharing (TBRS) theory (Barrouillet et al., 2004), which combines decay and interference theories and, thus, allows information to be "refreshed" as well. Still, recent TBRS implementations (Oberauer and Lewandowsky, 2008, 2011), exhibit rather low memory performance in terms of recall accuracy. Moreover, Adaptive Resonance Theory networks (Carpenter and Grossberg, 1987) are also used to encode memories. However, current implementations fail to effectively model a human-inspired long-term robotic memory, either because of limitations on the perceived information (Tscherepanow et al., 2012), or on the information retrieval and refreshing (Taylor et al., 2009) or, even, because of absence of forgetting mechanisms (Leconte et al., 2016).

Evidently, the exploitation of the stored information in order to infer or predict the state(s) of the interaction would greatly facilitate sHRI (Maniadakis et al., 2007). Yet, only recently there have been some works researching memory-based inferencing. For example, Hidden Markov Models (HMMs) are used in order to infer actions consisting of a sequence of "intentions" (Kelley et al., 2008). However, the need of previously modeled and task-dependent actions, limits the employment of the system in complex real-world setups. This obstacle is alleviated in some of contemporary works, such as Nuxoll and Laird (2012) and Petit et al. (2016). The former refers to the employment of the Soar cognitive architecture in order to exploit episodic memories and enhance the cognitive capabilities of artificial systems, while, the latter, uses a-posteriori reasoning to store and manage previously acquired knowledge. However, both of these works face severe limitations regarding the constantly increasing storage requirements, negatively affecting performance in long-term HRI.

To address the aforementioned issues, we have implemented a time-aware episodic memory module (Sigalas et al., 2017a,b) for autonomous artificial agents, which enables memory storage and management, as well as sHRI state prediction and inference. As thoroughly described in section 4.1, symbolic information is stored in a temporally rich domain, which encodes the involved entities and the relation between them. Each entity is characterized by an importance factor which dictates its life-cycle and, thus, determines whether to keep or erase the related information. Separate HMMs are generated and trained on demand in order to categorize the stored information, query the memory about past events, infer "hidden" information about an episode's attributes and predict future actions.

2.2. Temporal Information During Action Observation

Time perception, i.e., the ability to perceive the temporal properties of an ongoing activity, is a field that remains relatively unexplored in artificial cognitive systems. This can be attributed to the fact that such investigations often require task dependent and contextual data, which are difficult to obtain. Recently, however, temporal information has been increasingly used for action recognition, which indicates that there is a strong correlation between low-level behaviors and temporal properties. For example, local spatio-temporal features (Laptev, 2005) have been showcased to have increased discriminative power (Wang et al., 2009), since strong variations in the data (such as characteristic shifts in motion and space), can be captured more accurately in the spatio-temporal domain. In this context, various descriptors and feature detectors have been proposed. In Laptev (2005), temporal information is attained by convolving a spatio-temporal function with a Gaussian kernel, while in Scovanner et al. (2007), a 3D SIFT descriptor is proposed, which extends SIFT to the time-domain. Dollàr et al. proposed the Cuboid detector, which applies Gabor filters along the temporal dimension. Temporal structure for activity recognition has also been investigated using graphical models, including spatio-temporal graphs (Lin et al., 2009) and semi-latent topic models (Wang and Mori, 2009). In contrast to the aforementioned works, which examine the temporal structure of an activity, in the current work we measure its duration. As we demonstrate the use of this information can lead to robust descriptors for the human activity.

To facilitate temporal predictions by mere observation, we have introduced Generative Time Models (GTMs) (Hourdakis and Trahanias, 2018) that can accurately predict the duration of an unfolding activity. i.e., observation models that provide in real-time estimations of temporal quantities that characterize the activity. This concept, that is predicting the time-related properties of an activity, is novel to robotics and with great potential. Information provided by the GTMs can be employed by different disciplines including human-robot interaction, scene perception, robot planning, and process modeling. In the current work, GTMs are employed to implement the observation models that allow the robot to predict the duration and completion-time of an activity performed by a human agent.

2.3. Time-Informed Planning of Collaborative Activities

Several works have considered the notion of time in planning solo robot behavior in the form of action sequences, frequently with the use of PDDL that uses first-order predicates to describe plan transitions (Cashmore et al., 2015), or NDDL that considers a "timeline" representation to describe sequences of temporal conditions and desired values for given state variables (Py et al., 2010), and is adopted by the EUROPA Planning Library (Rajan et al., 2007; Barreiro et al., 2012) and its subsequent advancement that considers the description of hierarchical plans (Antonelli et al., 2001). Opportunistic planning provides an alternative view for scheduling long-horizon action sequences (Cashmore et al., 2018). The use of hierarchical plans is additionally considered in Stock et al. (2015), focusing on the unification of sub-plans to improve implementation efficiency. Moreover, the high-level Timeline-based Representation Framework provides a structured library for managing operational modes and the synchronization among events (Cesta et al., 2009), or with the use of the forward search temporal planner POPF (Cashmore et al., 2014). Extensions of this framework has been used among other in industrial human-robot collaboration (Pellegrinelli et al., 2017; Umbrico et al., 2017) to ensure controllability.

To implement tasks involving multi-agent collaboration, planning algorithms often rely on constraints which provide ordering between the independently implemented activities (Morris et al., 2001; Shah et al., 2007; Smith et al., 2007; Morris, 2014). Existing approaches explore the controllability of alternative strategies, to identify plans that successfully schedule the required activities in a way that satisfies constraints until the final completion of the goal (Cimatti et al., 2016). Despite their success in coordinating pairs of interacting agents, relevant works suffer in terms of scalability because they assume a significant amount of resources to be devoted to the formulation of the full plan.

Interestingly, relevant works consider the use of time in full isolation, without the ability to blend time with other quantities for the time-inclusive multi-criteria evaluation of plans. For example, time-labeled Petri-nets have been used to accomplish fluent resource management and turn-taking in human-robot collaboration focusing mainly on dyadic teams (Chao and Thomaz, 2016). In a different work time has been sequentially combined with space to minimize annoyance among participating agents (Gombolay et al., 2013). Other works follow a similar tie-isolated formulation, representing agent interaction as a multi-criteria optimization problem. The objective function is derived from the preference values of participating agents and the temporal relations between entities are mapped on the constraints of the problem (Wilcox et al., 2013). More recent works follow basically the same formulation, representing time in the set of constraints that confine available solutions (Gombolay et al., 2017). Besides the fact that criteria such as the workload and the user preferences can be addressed with these approaches, time is largely kept separate form other quantities, thus not used for the formulation of time-informed multi-criteria objectives. Moreover, the works mentioned above do not consider predictive estimates on the performance of interacting agents and the expected release of constraints among tasks.

Recently, decentralized approaches are used for multi-robot coordination, which work on the basis of auctions. For example, Melvin et al. (2007) considers scenarios in which tasks have to be completed within a specified time window, but without allowing overlapping between time windows. Modern approaches are targeting this issue with particularly successful results in simulation environments (Nunes et al., 2012; Nunes and Gini, 2015). In other similar problems the routing of working parts is assigned to the most suitable transportation agent through an auction-based mechanism associated to a multi-objective function (Carpanzano et al., 2016). However, the relevant approaches assume auctions to proceed on an agent-centered point of view which does not consider the capacities and special skills of other team members. Therefore, it is hard to maximize the usability of all members for the benefit of the team (i.e., it might be beneficial for the team if the second optimal agent undertakes a given task).

To address the issues mentioned above, we have recently introduced the Daisy Planner (DP) (Maniadakis et al., 2016a,b), a new scheme of time-informed planning, which relies on the daisy representation of tasks and adopts time-inclusive multi-criteria ranking of alternative plans. DP operates under the assumption of pursuing immediate, locally optimal assignment of tasks to agents. This is in contrast to previous works on scheduling multi-agent interaction that typically prepare long plans of agents' activities for all future moments (Gombolay et al., 2013; Hunsberger, 2014; Cimatti et al., 2016), under the risk of frequent re-scheduling, due to external disturbances that may render current plans inapplicable. In such cases, re-scheduling may take up to a few tenths of seconds (Pellegrinelli et al., 2017). DP effectively operates as a lightweight process which minimizes the chances for re-planning in the case of unexpected events (Isaacson et al., 2019).

3. MULTI-AGENT INTERACTION SCENARIO

Without loss of generality, and for the sake of brevity, the adopted scenario considers the case of three agents: a human and two robots (a humanoid and a robotic arm). We note, however, that the proposed methodology is readily applicable to the case of more than three cooperating agents. In this section, we summarize the scenario that will be used as a motivating example for the rest of the paper (see **Figure 1**). The scenario assumes that the three agents cooperate for the timely delivery of breakfast to the human.

In particular, to predict the breakfast preferences of the human at a given day, the system exploits past sHRI sessions. After querying the episodic memory based on the current date (i.e., day of week, month, and season), the weather conditions and the user's mood and health, the most likely breakfast menu is inferred and forwarded to the planner that synchronizes the activities of involved agents. The collaboration scenario assumes agents to undertake the tasks they are more efficient to implement. Therefore, the human undertakes table cleaning for the breakfast to be served, given his superior performance for the task in comparison to the two robots. At the same time, the humanoid robot gets the responsibility to fetch the breakfast to the table. We use a bowl carefully mounted on humanoid's chest to support the transfer of breakfast items. This is accomplished with the help of the robotic arm, which places the appropriate number of items in the humanoid's bowl.

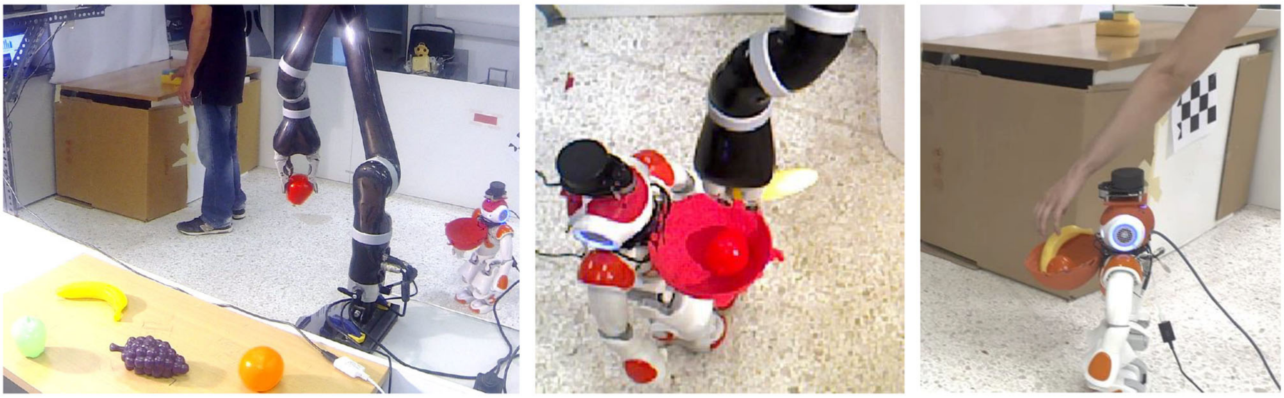


FIGURE 1 | The human-robot-robot "breakfast preparation" collaborative evaluation scenario.

The number of breakfast items is dynamically determined according to the progress of table cleaning by the human. The collaboration of the two robots aims at getting and fetching the maximum number of items, after considering the performance profile of the human and additionally that the breakfast should be delivered with the minimum possible delay. As a result, a fast performing, dedicated human that is willing to finish breakfast as soon as possible may be served a minimal breakfast menu, while a very relaxed person will, on the contrary, enjoy a full breakfast menu. The above described breakfast delivery scenario facilitates the grounded assessment of the integrated time-informed system which accomplishes the fluent and timely synchronization of the robots with human activities.

3.1. Task Requirements and System Modules

The implementation of the above described scenario assumes the integrated performance of modules targeting diverse parts of the sHRI timeline. More specifically, to capitalize on the information gained from past sHRI sessions, the composite system must be able to (a) maintain a temporally-rich representation of past HRI events being easily searched using temporal criteria, which enables focusing on the past time periods of interest and (b) accurately infer or predict the state of the HRI, based on past experiences.

Beyond associating relevant past experience to the current situation, fluent HRI requires real-time monitoring of human activities. Our implementation achieves this (a) by developing accurate predictions on the completion of human actions with few training iterations and minimal prior information about the performed activity, and (b) by estimating human efficiency toward the real-time profiling of the interacting human.

Following the above, it is important to effectively proceed to the accomplishment of the joint goal, by coordinating the activities of team members. To this end, the multi-agent action planner is necessary to (a) maximally exploit the individual skills that each one of the heterogeneous participating agents brings into the team, and (b) implement plans that are flexibly and proactively adapted to the expectations of the user and the dynamic temporal characteristics of the collaborative scenario.

4. TIME-INFORMED MULTI-AGENT INTERACTION IMPLEMENTATION

Figure 2 depicts the sHRI composite system, designed according to the previously delineated requirements, featuring the interconnections of the individual time-aware cognitive modules. The implemented modules are involved in distinct aspects of the sHRI process extracting complementary pieces of temporal information. As evident in **Figure 2**, the Episodic Memory builds on past experiences to make predictions of the human agents' needs and preferences. The relevant predictions are fed to DP which guides and coordinates agents' activities toward the implementation of the mutual goal, adequately synchronized with the evolution of real-world events. The GTM module effectively monitors the progress of human task implementation to predict remaining time and user efficiency, which are used by the DP to successfully steer and refine the cooperative plans. The extracted temporal features are additionally encoded to memory for future reference. The above described continuous interplay of Episodic Memory, GTM, and DP results into a composite system with a context and human personality driven performance that accomplishes to effectively map robot services to the needs of the individual humans.

The proposed approach implicitly addresses issues regarding the commitment of the agents to their common goal (Castelfranchi, 1995). This comes from the central coordination of the team by the planner, which eliminates motivational, instrumental, and common ground uncertainty as they are described in Michael and Pacherie (2015). In that sense, every member of the team holds normative expectation from the others, which are crucial for the successful accomplishment of the common goal (Castro et al., 2019). However besides the coordination of the team by the planner, currently, there is no means to explicitly communicate expectations or obligations among partners, an information that might be crucial for the

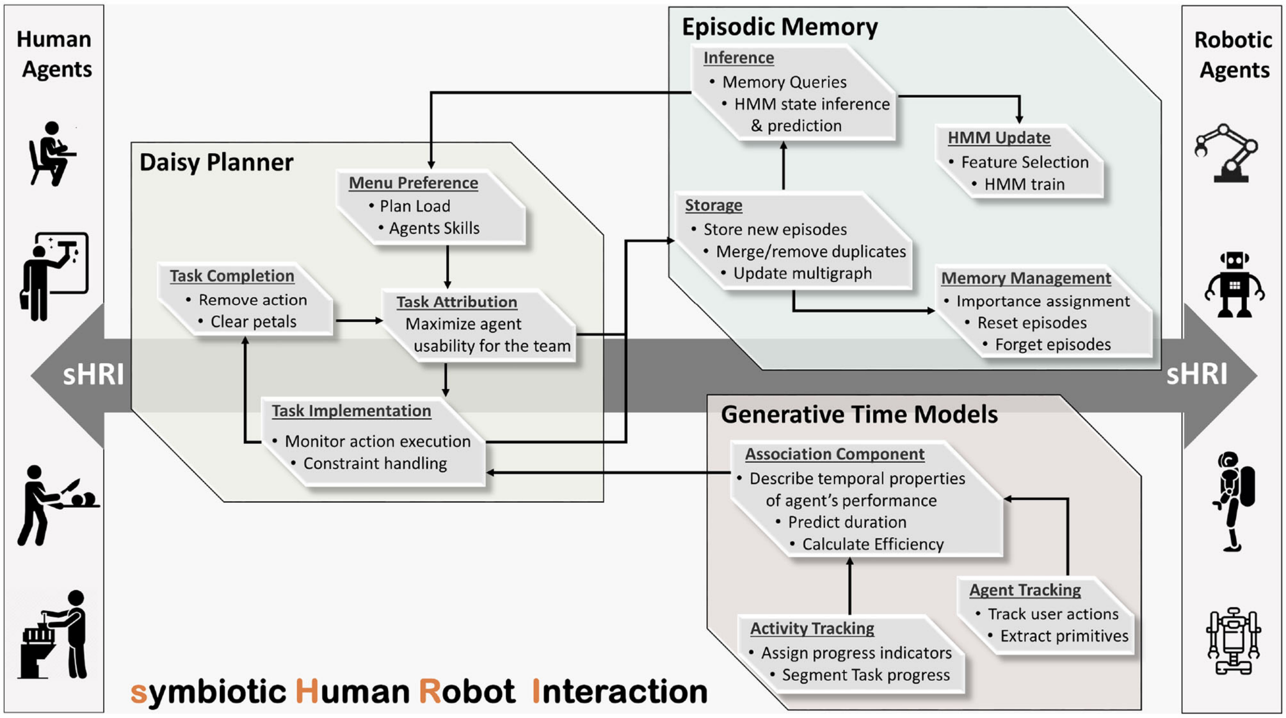


FIGURE 2 | Abstract representation of the proposed sHRI system.

human to understand that things are under control and the task progresses as expected. This is something that will be considered in the future versions of the system.

4.1. Past–Encoding of Elapsed HRI Sessions

Development of an episodic memory module, able to effectively encode past episodes on a temporally rich domain, may significantly facilitate sHRI by exploiting past experiences to infer current human needs. Specifically, the episodic memory serves to store, manage and symbolically represent user memories, in a manner which, on one hand enables storing of large numbers of entities and, on the other hand, facilitates fast and efficient search. Typically, memory stores and manages all of the perceived information. However, for the task at hand, we focus on the users' breakfast preferences, storing, and exploiting only breakfast-related entities along with their temporal (e.g., date of occurrence, duration of activity, etc.) or other (user mood and health, weather conditions, etc.) information.

Turning to the usability of episodic memory, robotic systems should ideally adapt their activities in accordance to user needs. To this end, exploitation of elapsed sHRI episodes may significantly facilitate the inference of user preferences at the given context. Valuable information stored in memory may regard the configuration of past breakfast menus (i.e., combinations of breakfast items) in association with the evolution of relevant attributes, e.g., human mood and health, on a daily basis.

4.1.1. Episodic Memory Design

As described in our previous works (Sigalas et al., 2017a,b), episodes are stored in memory in the form of connected multi-graphs (**Figure 3A**), where the nodes represent the episode entities (i.e., Scenario, Event, Action, Actor, Object, or Feature) and the edges represent the links (bi-directional parent-child relationships) among entities, during the unfolding of the episode. Each link of the multi-graphs is assigned an importance factor which affects the entity's lifecycle and varies according to the ongoing task, i.e., entities which are more "relevant" to the task are considered more important than other "irrelevant" ones. Importance is represented by a damped sine wave (i.e., decays over time), as shown in **Figure 3B**. The latter is mathematically formulated as:

$$I(t) = e^{-\lambda t} \cos(\omega t). \tag{1}$$

where λ is the amplitude decay constant, ω is the angular frequency and t represents the lifetime of the entity ($t = 0$ at the first occurrence of the said entity).

The information stored in memory is evaluated and updated periodically (at the positive peaks of the importance sinusoidal) where it can be either refreshed, merged with an overlapping entity, or forgotten, as summarized below.

- Refresh: Each time an entity is perceived or considered by the system, its importance timeline is set to $t = 0$, which indicates that it is refreshed as a memory-entity, and its importance starts decaying again following Equation (1).

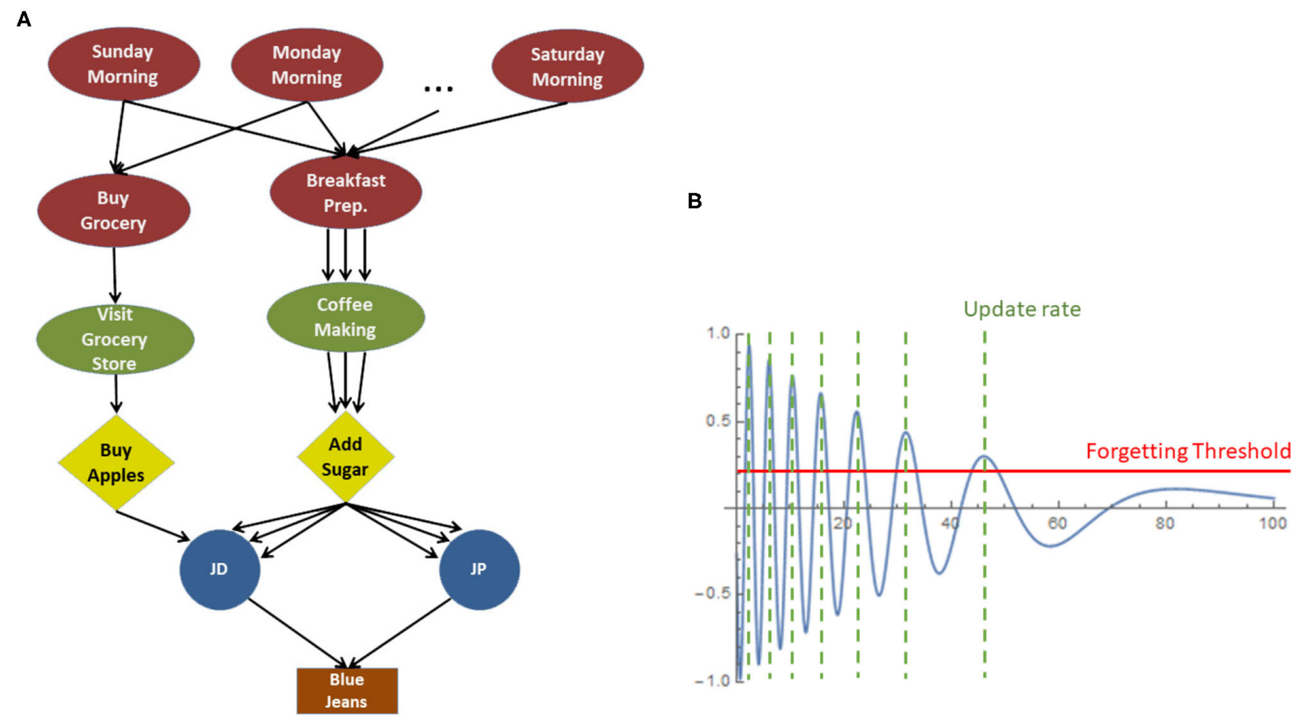


FIGURE 3 | (A) Connected Episodic Multi-graph. **(B)** Exemplar importance function.

- Merging: Repeated instances of the same entity (e.g., a person seen twice) are merged together. This is feasible due to the multigraph nature of the memory, allowing for multiple "incoming" and "outgoing" links, as illustrated in **Figure 3A**. The multigraph provides efficient memory indexing and thus facilitates fast response on memory recall queries. The given representation allows for queries of the type "*what happens every morning*" or "*what happens every time user JD is sick*," and thus facilitates statistical analysis on human behavior and preferences.
- Forgetting: The efficient management of memory assumes a forgetting mechanism to filter out "noise," which means to discard entities being of low importance for the stored episodes. Whenever an entity's importance drops below the so called "forgetting threshold," the entity is deleted from the memory together with all adjacent (incoming and outgoing) association links with other entities. *Forgetting* is an iterative process, in the sense that, erasing an entity affects also its children, which, if left with no incoming links (i.e., have no other "parent" than the erased one), will be erased as well.

4.1.2. Probabilistic Inference

In order to exploit the stored information and make predictions about user needs and preferences, we employ a Hidden Markov Model (HMM) inference schema (Sigalas et al., 2017a). By querying the memory, it is possible to retrieve information about past episodes, properly filtered by selection criteria. These criteria vary depending on the task at hand and the required inference; e.g., "what does JD eats for breakfast during weekdays," or "what did JD say when the phone rang yesterday morning."

Separate HMMs are developed on the fly—and on demand—to exploit the time-stamped data retrieved from memory. The recalled past episodes along with the selected attributes, are used to train the HMM (estimate its parameters) and infer scenario-relevant information. Training is accomplished by employing the forward-backward algorithm (Rabiner, 1989), a two-step iterative process that uses observations to predict the model state, which is subsequently used to update the model parameters. Similarly, in order to make a history-based inference of a user's preference, the HMM exploits the observed episode attributes to predict the upcoming state, based on the currently estimated model parameters. To facilitate training, a feature selection mechanism (for the current implementation we use the Boruta algorithm; Kursa et al., 2010) is periodically employed, in order to select the most relevant—to the query—features and, thus, increase inference accuracy.

HMMs are perfectly suited for the task at hand, because they provide a very flexible generalization of sequence profiles allowing for inputs of varying length. Moreover, they efficiently encapsulate the hierarchical structure of real world episodes while they are also incrementally trained, allowing for fast operation during the online scenario unfolding.

By exploiting stored information in combination with the HMM-based inference, the Episodic Memory module manages to: (a) estimate the HRI state, e.g., agent actions in relation to the objects in the scene, (b) infer hidden HRI information, e.g., user's intentions, and (c) identify abnormal unfolding, e.g., emergency events or unexpected situations. In the scenario considered in the present work, the HMM is used to predict the breakfast preferences of the user, which are further fed to the planner in order to effectively guide the scenario unfolding.

4.1.3. Episodic Memory Enhancement

The Episodic Memory module was further enhanced toward the direction of increasing the performance of the HMM-based inference mechanism, in terms of both the efficiency (i.e., high inference/prediction accuracy) and effectiveness (i.e., robust and fast HMM training and inferencing). To this end, we extended the inference by: (a) not discarding the HMM after usage and (b) periodically train each HMM, instead of updating it only when queried. Initially, a feature selection mechanism [the Boruta algorithm (Kursa et al., 2010) as already mentioned above] is employed to select the most relevant -to the task- features, which are then used to incrementally train the HMM.

4.1.4. Episodic Memory Key Strengths

The above presentation dictates that, in comparison to previous relevant works, the Episodic Memory module bears important features:

- Encodes episodes as symbolic information on a temporally rich domain.
- Dynamically manages (e.g., merges or forgets) the stored episode details, based on their temporally decaying importance.
- Provides accurate inference about the current or future state(s) of the HRI, based on the personalized preferences, as derived from the stored information.

4.1.5. Episodic Memory Interface

The episodic memory is the representation of user's past experiences. It is directly interfaced with the Daisy Planner either for storing new information, or for inferencing the current of future state(s) of the unfolding scenario. The interface and the capacities that the memory module brings to the system, are summarized below.

Input. The memory accepts two types of input from DP. (a) Whenever an action is accomplished, the planner feeds memory with the relevant information; i.e., involved entities and their characteristics, general information about the current day and so forth; (b) DP sends requests about the ongoing or forthcoming states of the HRI. These requests are formulated as plain database queries, stating the predicted value(s) and the accompanying constraints.

Output. The output of the inference mechanism depends strongly on the incoming query. Based on a given request and constraints, the HMM is updated accordingly and the most likely response is fed to the planner.

Role. As evident, the episodic memory module serves two purposes: Storing and managing of past episodes, using a time-aware symbolic representation and estimating the current or future state(s) of the ongoing scenario, based on the time-stamped information and the corresponding (temporal or other) constraints.

4.2. Present—Temporal Features of Perceived Human Activity

Temporal information, i.e., activity duration, allows robotic systems to plan their actions ahead, and hence allocate effort and resources to tasks that are time-constrained or critical. In human cognition, such perception models are widely used (Zakay, 1989), despite the fact that our time-perception is subjective, and dependent on the implicit *sense-of-time* feeling that stems from our sensorimotor pathways (Zakay and Block, 1996). In contrast to that, robots and artificial systems may potentially perform this task more consistently, by observing and analyzing the statistical properties of the observed behaviors (Bakker et al., 2003). Recently, we have demonstrated how such duration estimates can be obtained using a model based method to derive the progress of the activity (Hourdakis and Trahanias, 2013), called Generative Time Models (GTMs) (Hourdakis and Trahanias, 2018).

4.2.1. GTM Design

For the current implementation, GTMs are used to observe, analyze and subsequently predict the temporal properties of the human's activity (see **Figure 4**). This is accomplished by segmenting and decomposing the observed activity based on the human's motions. For the example of wiping the table, where we have repeating oscillatory motions, the primitives are described by their amplitude and period. To obtain the primitives, a GTM segments the signal obtained by tracking the human hand, by looking for local extrema at small Δt intervals and stores their starting t_s and ending t_e times. To evaluate the local extrema it looks into the derivative of the signal, which at a point of a peak has a zero-crossing at the peak maximum.

To identify peak signal positions, the algorithm smooths the signal's first derivative, by convolving it with a Gaussian kernel, and then stores the indices of the zero-crossings on the smoothed derivative. For each index, a prominence value is calculated, which indicates whether there has been a significant change in the motion direction vector. The algorithm returns the n largest peaks whose prominence exceeds a certain threshold value.

The current work focuses mainly on the table wiping task, using a GTM to extract the oscillatory motions that the human produces, and associate them with the task progress (**Figure 4**). However, the GTM concept can easily generalize, because it uses a modular architecture, with the activity and agent components kept separate. Consequently, components from a GTM formulation can be re-used to other tasks. To create robust temporal predictors, a GTM analyzes an activity using two observation models: (1) task progress, and (2) control. The first estimates the progress of the task, i.e., how much of the activity has been completed. The latter, identifies and records information about the observed motions that appear during the activity. For each motion, it records (i) the effect it has on the task progress, i.e., how much of the task is completed each time the motion is executed, (ii) the duration of each motion, and (iii) how frequently it appears during the activity. A GTM uses this

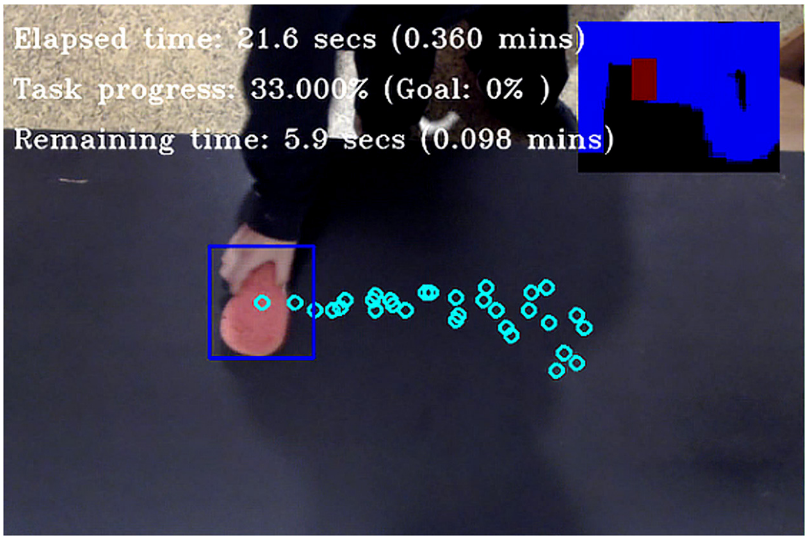


FIGURE 4 | A GTM observing a human, while cleaning a table. The model extracts the low-level behavioral primitives of the human and associates them with the quantified task progress (small top-right plot). For the "wipe the table" activity, task progress is calculated by measuring the percent of the table surface that has been wiped by the sponge.

information to predict the probability of a motion occurring, i.e., how many times each motion is expected to appear in the future in the course of this activity, and how it will affect its progress.

To accomplish this, a GTM builds behavioral profiles based on motion primitives that are observed, and uses them to infer future task progression in respect to the human performance. To create its temporal predictors, the GTM employs the observed primitive models. For each primitive observed, the model segments the overall motion, and uses those segmentation intervals in order to infer how the task progresses in each interval. To make the predictions, the model follows a finite mixture approach, in which a belief is formed about the probable primitive models that will be observed by the model.

Having segmented and described each primitive that is observed, a GTM approximates the activity progress O in future time-steps, using a finite mixture model. To estimate O we sum the expected progress to the task by each observed primitive, weighted by the primitive's probability, as shown in Equation (2):

$$O(t) = \sum_{i=1}^{k} p(i)\phi_i \tag{2}$$

where $p(i)$ are the weight factors, that satisfy $p(i) \geq 0$, for all integers $i \in 1, k$, and $\sum_{i=1}^{k} p(i) = 1$, while $\phi_i = \int_0^t f_{M_{(i)}}(t)dt$ provides the overall contribution of the primitive i to the task progress, with $f_{M_{(i)}}(t)$ being the function that describes how each primitive contributes to the task progress at a certain point in time. A GTM uses Equation (2) to predict future states for the activity progress, i.e., the expected change for the task progress is calculated using the probability of observing the primitive, and how much the latter contributes for the task completion. Using Equation (2), one can derive useful information about the observed activity. Given the weight factor $p(i), \forall i \in 1, k$ for all primitives one can estimate, using Equation (2), how the task progress will change from t to $t_k = \sum_{i=1}^{k}\left(p(i)d_i\right)$:

$$O(t + t_k) = O(t) + \sum_{i=1}^{k}\left(p(i)\int_{t}^{t+d_i} f_{M_{(i)}}(t)dt\right) \tag{3}$$

Equation (3) provides an estimate of the activity progress forward in time, using the $f_{M_{(i)}}$ as basis functions. Based on Equation (3), robust predictions on the duration of an activity can be obtained. For the current implementation, the model is used to provide estimates that can infer how long a human agent will require in order to finish the table wiping task. A more detailed presentation of the above model can be found in Hourdakis and Trahanias (2018).

4.2.2. GTM Enhancement

For the current implementation, GTMs were extended to estimate the efficiency of the agents when performing a task, i.e., the extent to which the actions performed are productive toward finishing the activity. To this end, efficiency is relevant to self-learning, and measures the quality of task execution for a given activity. To accomplish this, we measure for each primitive the fraction of the percent of the activity it completes against its duration (Equation 4).

$$e_h = \frac{Pr_t}{Pr_d} \tag{4}$$

where Pr_t indicates the percent of the task that has progressed due to a primitive, and Pr_d the duration of that primitive. Both quantities are readily available and computed using the GTM mathematics, as described in Hourdakis and Trahanias (2018).

Efficiency values are estimated online while the human activity progress, and they are sent to the planner for further processing and timely adaptation of the multi-agent collaboration strategy.

4.2.3. Key Strengths

GTMs can make accurate predictions online, making them an ideal candidate to process the immediate planning context of an interaction session. Their key strengths are summarized below:

- Provide robust predictions with few training iterations.
- Use a modular architecture, with segregated Control and Activity observation components, which allows the concept of GTMs to generalize across tasks.
- Can extract additional metrics, such as efficiency, which are useful for planning.

4.2.4. GTM Interface

GTMs provide a module that profiles and predicts the future performance indicators of a human. Below we outline the module's input, output conventions and role in the composite system.

Input. Input in the GTM is in the form of visual images, obtained by a camera. For the current experiment, the camera is mounted approximately 2 m above the table, in order to observe the table wiping task. At initialization, the human marks the rectangle containing its hand, which is used by the GTM for tracking.

Output. Using the raw camera images and tracker input, the GTM identifies the primitives of the human, and estimates two measures: (1) the expected duration of the experiment, and (2) the human's efficiency. This information is subsequently sent to the planner.

Role. The role of the GTM is to extract and estimate a temporal profile of the human participating in an interaction session. This profile is used to predict future task states, and temporal parameters regarding the human's performance.

4.3. Future—Plan Robot Behaviors in Coordination With Human Activities

The fluent coordination of multi-agent activities plays a crucial role in the joint accomplishment of goals. We have recently introduced (Maniadakis et al., 2016a,b), a time-informed planner that attributes tasks to agents in a step-by-step manner, accomplishing the effective coordination of multiple agents (see also, Isaacson et al., 2019). The planner assumes the daisy-like representation of the composite behavior and is thus termed *Daisy Planner* (DP). In particular, each task consisting of an action sequence is represented as a petal of the composite daisy graph. Constraints link actions among tasks that can be implemented in parallel, to indicate that the completion of a certain action is a prerequisite for the action of the other task to commence.

The planer is designed as a lightweight immediate optimal planning module, particularly appropriate for dynamic multi-agent environments where unexpected events (e.g., a phone ring, or the drop of human performance) may increase the implementation time of tasks and trim off team productivity. The DP avoids searching extended solutions of complex agent-task assignments that span over the future timeline, in order to flexibly and with low-cost adapt to unpredicted circumstances. The local view of the planner makes processing particularly light-weight, because it does not synthesize and does not compare complex future scenarios as it is the case with previous works (Wilcox et al., 2013; Gombolay et al., 2017), which additionally suffer from the need of resource-expensive rescheduling when unexpected events occur.

The planner functions under the assumption of task assignment to agents based on their availability. In order to find the best petal fit for a given non-busy agent, DP considers the capacities of all team members and builds upon the skills that the current agent brings into the team, trying to make it maximally useful for the team and the given interactive scenario. This is different to existing approaches based on Timed Petri Nets (Chao and Thomaz, 2016) in which agents are pre-assigned the sets of actions they are implementing. The planner effectively combines time with other quantitative measures that outline key features of task implementation, such as efficiency, robustness, even fatigue, and like/dislike for the case of humans, in order to construct composite time-inclusive criteria for ranking alternative multi-agent plans. This is in contrast to other works that include time as a constraint that confines the search of viable solutions (Gombolay et al., 2017).

4.3.1. DP Design

The setup of the DP assumes the identification of tasks that have to be fully implemented by a single agent. For example, to implement the task "pour oil in salad," the very same agent must grasp the oil bottle, move it above the salad, pour the oil and put the bottle back on the table. Therefore, "pour oil in salad" is represented as a petal of the composite daisy-represented scenario. Each task/petal consists of a sequence of actions that start and end at the rest state.

To initialize DP, the duration and quality of implementation for all possible action-agent pairs is provided to the planner. Duration information is obtained by summarizing previous trials and has the form of (min, max) experienced time. The quality of implementation is set by the experimenter, e.g., the humanoid is declared with poor quality to grasp and manipulate complex objects, but high quality to navigate. Using this information, the DP successfully matches tasks with the skills of individual agents, being able to construct particularly productive teams, which may flexibly consist of heterogeneous agents.

The planner employs the fuzzy number representation of time to facilitate the processing of temporal information (Maniadakis and Surmann, 1999). Following the well-known representation of fuzzy numbers in trapezoidal form with the quadruplet (p, m, n, q), a fuzzy duration in the form "approximately a to b moments" is represented with the fuzzy trapezoidal number $(0.9a, a, b, 1.1b)$. In the current work, parameters a and b correspond to the minimum and maximum experienced implementation times, as discussed above. The use of fuzzy calculus (Dubois and Prade, 1988) provides the means to effectively associate the temporal properties of individual actions, predict delays of alternative

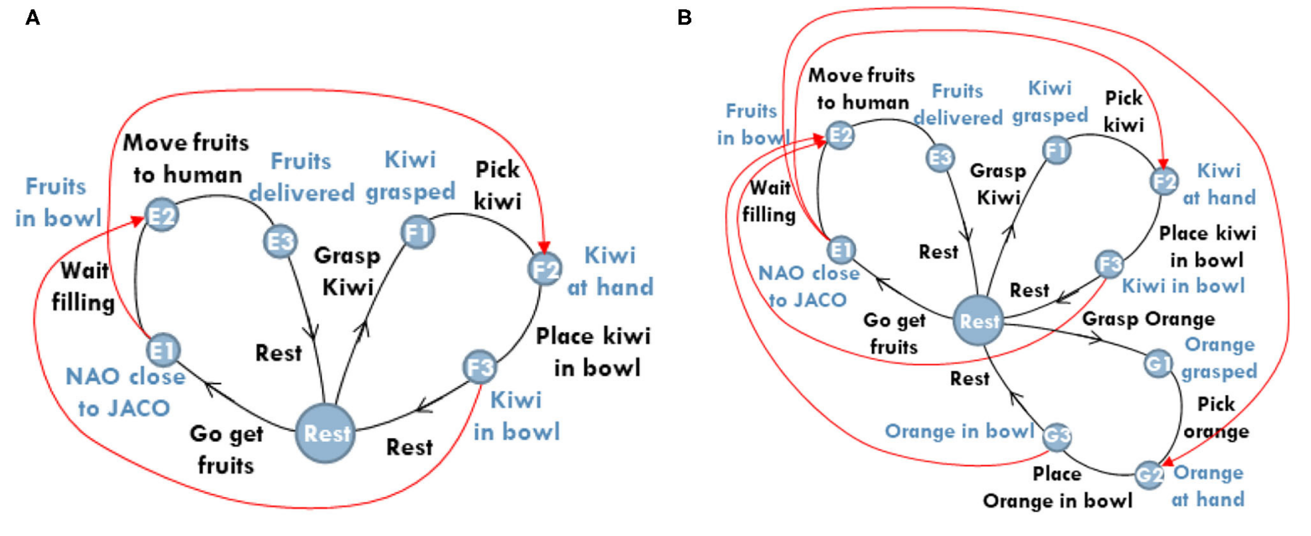


FIGURE 5 | Exemplar cases of plan constraints, depicted in red. **(A)** Illustrates the case of two constraints used to prioritize the actions of petals implemented in parallel. **(B)** Illustrates the case of two constraints on *E*2. As soon as *F*3 is reached, the planner releases the constraint *G*3 → *E*2, to comply with temporal constraints (i.e., avoid the delay of breakfast delivery).

planning scenarios and enable corrective measures to be taken in order to enforce the coordination of the individual activities (Maniadakis and Trahanias, 2016a). Moreover, it facilitates the comparison of agents' utility on different tasks (by combining implementation time and effectiveness), therefore enabling the use of optimization criteria for the locally optimal attribution of tasks to agents. In particular, each non-busy agent is assigned a new task in a way that maximizes agent's utility for the team, given the current, short-term view of team performance. Full implementation and assessment details of DP have been presented in Maniadakis et al. (2016a,b), and are not repeated here and re not listed here for the clarity of presentation.

In short, the immediate optimal planning approach followed by the DP, aims at naturalistic, smooth and low anxiety collaboration among the participants rather than generating globally-optimized minimum-time behaviors. This is particularly the case in most human daily collaborative tasks where participants share jobs based on expertise, tiredness, etc.

4.3.2. DP Enhancement

The current article elaborates on the management of constraints which prioritize action execution between tasks that may implement in parallel but constrain each other. In particular, the present work considers the time each constraint is expected to release in order to make more informative decisions when attributing agents to tasks.

For example, consider the case shown **Figure 5A**, where two constraints (shown in red) determine the ordering of action execution between agents working on different tasks. The first constraint (top red arrow) specifies that the humanoid robot must have completed its way to the robotic arm, before the latter starts placing the fruits in the humanoid's bowl. The second constraint (bottom red arrow) specifies that the fruit should be in the bowl before the humanoid departs to deliver breakfast to the human.

To effectively manage time resources, the planner needs to know when the humanoid is expected to arrive close to the robotic arm. The planner knows that navigation was initiated, for example, 34 s ago and the whole navigation takes approximately 50–60 s, represented by the fuzzy number quadruplet $(45, 50, 60, 66)$. Therefore, the remaining time for humanoid's navigation is $t_{hr,n} = (45, 50, 60, 66) - 34 = (11, 16, 26, 32)$. At the same time, the time needed by the robotic arm for grasping the fruits is known from previous trials to be $t_{ar,g} = (10.8, 12, 20, 22)$ and for picking the fruits $t_{ar,p} = (2.7, 3, 4, 4.4)$. Thus, the total time needed by the arm to prepare fruit placement is $t_{ar,g+p} = (13.5, 15, 24, 26.4)$. The difference[1] between $t_{hr,n}$ and $t_{ar,g+p}$ according to the LR-calculus, results to the fuzzy number $(-15.4, -8, 11, 18.5)$ which reflects the predicted desynchronization of the two agents. The defuzzification of this interval (implemented by the classic graded mean integration representation; Khadar et al., 2013) results into 1.516 s, indicating that robotic arm is not yet delayed, but should soon proactively initiate fruit grasping to avoid introducing idle time in humanoid's schedule. It is noted that the forward looking, proactive release of constraints based on the real-time monitoring of scenario unfolding is a new feature that has not been addressed by previous works.

Moreover, in the current work DP is enhanced to develop personalized sHRI that exploits real-time human temporal profiling, thus introducing an additional new feature to the state of the art. To slightly complicate the scenario considered in the present study, we assume that a high performing human might

---

[1] The subtraction of fuzzy numbers is as follows: $(a_1, a_2, a_3, a_4) - (b_1, b_2, b_3, b_4) = (a_1 - b_4, a_2 - b_3, a_3 - b_2, a_4 - b_1)$.

probably be in a high arousal state, he/she highly dislike delays, and would only be satisfied with the delivery of the breakfast immediately after table cleaning. In contrast, a low performing person may be in a low arousal state and most likely would thoughtlessly accept small delays in the order of a few seconds, with the benefit of having more fruits delivered. The above builds on the well studied link of emotional state and time perception, which shortly claims that time seems to fly when we are in a high arousal state, and to drag on when we are bored (Droit-Volet and Meck, 2007).

Following the scenario, human efficiency $e_h \in [0, 1]$, defined in Equation (4), is an important parameter for determining the number of breakfast items to be served to the human. To explain this further, we consider a second example focused on multi-agent collaboration (see **Figure 5B**). We assume that the planner is informed of the estimated remaining time for the human to complete the cleaning task $t_{h,cl}$ and his current level of efficiency e_h. The robotic arm has just added an item in humanoid's bowl and the planner is ready to decide whether there is enough time for the arm to add one more item in the bowl, or, the humanoid should start navigation toward the human, to avoid delay. The planner knows by experience that the total time required by the arm to grasp, pick and place an item in humanoid's bowl is $t_{ar,g+p+p}$ and additionally that the time needed by the humanoid to deliver breakfast to the human is $t_{hr,d}$. The sign of the defuzzified difference between the total robot synergy time and the human time scaled by his/her efficiency is used to decide task allocation as described below:

```
$robotT = t_{ar,g+p+p} + t_{hr,d}$;
$humanT = \frac{1}{0.5+e_h} t_{h,cl}$;
if $defuz(robotT - humanT) < 0$ then
    Arm adds a new item in humanoid's bowl;
else
    Humanoid delivers breakfast to human;
end
```

Clearly $t_{h,cl}$ and e_h can drastically affect planner decisions. This is not only because decision making assimilates the latest estimate of human completion time $t_{h,cl}$, but additionally because efficiency values e_h are used to scale the human available time. In particular, e_h values close to one reduce further the human available time *humanT*, to stress the assumption that a highly efficient human in high arousal does not accept delay in breakfast delivery. On the other hand, e_h values close to zero have an opposite effect increasing the available time *humanT*, thus indicating that relaxed humans would tolerate a short delay under the benefit of having a richer breakfast.

Noticeably, besides the fact that previous works have also considered user efficiency in multi-agent interaction (Gombolay et al., 2017), the real-time assimilation of the relevant information to adapt team performance is an aspect that has not been addressed so far by existing works, but is greatly and inherently facilitated by the immediate planning approaches adopted by the DP.

It is noted that even if the planner decides the immediate depart of the humanoid, it may often be the case that active constraints inhibit the humanoid's navigation. This might be because not all three items are yet placed in the humanoid's bowl. This is particularly the case depicted in **Figure 5B**. However, given that the humanoid should preferably depart to avoid delay, the planner has the option to release all the constraints inhibiting the humanoid's departure, making the robot free and ready to go.

4.3.3. Key Strengths

In comparison to previous relevant works, the Daisy Planner:

- Implements plans that are flexibly and directly adapted to the dynamic unfolding of the collaborative scenario, which due to the immediate planning approach adopted, avoids re-planning of multi-agent activities.
- Operates as a lightweight process that effectively scales to handle large multi-agent teams, because the complexity of short-term task attribution increases linearly with the number of agents.
- Exploits the predicted temporal features derived from the real-time monitoring of agents' activities in order to enhance coordination between team members and more accurately meet the expectations of users.

4.3.4. DP Interface

The planner is the eye to the future for the composite system. It is directly interfaced with the Episodic Memory and GTM to develop and maintain a dynamic third-person perspective on user expectations. The interface and the capacities that the planner brings to the system, are summarized below.

Input. The planner sends queries to the memory to get back inferred estimates of the human preferences which are interpreted as the goal that the composite team has to achieve. To adequately orchestrate interaction, the planner is informed about the progress of action execution by the individual agents and the efficiency of human on the action he/she is currently implementing.

Output. The planner tracks the implementation of tasks by the individual agents and requests the timely execution of the relevant actions to enhance coordination. Additionally, it informs the memory about the evolution and the implementation details of the composite scenario (which agent implemented each task, when the implementation started and how long it took), which are stored for future reference.

Role. The DP actively guides the participating robots to map their activities on human expectations and times. It is implemented as a lightweight procedure that (i) composes effective multi-agent teams consisting of heterogeneous members, (ii) exploits information on the human behavioral profile to develop assumptions about his/her temporal expectations and accordingly adapt the performance of robotic agents, (ii) enforces the timely interaction among agents considering the inter-dependencies between the individual activities, (iii) provides to the system a third person perspective on how humans perceive the notion of time, and (iv) flexibly adapts robot activities to the performance of the other team members (e.g., robots may speed up to catch up a fast performing human).

5. EXPERIMENTAL RESULTS

The proposed approach has been implemented and validated in a realistic scenario that regards the interaction of three agents, i.e., one human and two robots, as summarized in section 3. In the current work two robots are used, namely the Kinova JACO six-joint robotic arm manipulator and the Softbank Robotics NAO humanoid, which contribute to the robot team complementary skills for serving the human. The details of technical implementations and the experimental setup are described below, followed by the real-world and the quantitative assessment of specific modules and the composite system as a whole.

5.1. Enabling Robotic Skills

To implement the scenario discussed throughout the paper in the real world, a variety of robotic modules have been implemented to facilitate task accomplishment by the individual agents and guarantee the success of the synergistic multi-agent performance.

5.1.1. NAO Mapping and Localization

Initially, a 2D-map of the environment is created utilizing a planar-LIDAR mounted on the NAO robot's head and the leg odometry with the ROS mapping package[2]. This map is a typical occupancy grid highlighting where obstacles are located. Subsequently, the robot can localize itself in the map with a particle filter fusing in real-time the laser scan readings and the leg odometry. This is done with the Adaptive Monte Carlo Localization ROS package[3].

5.1.2. NAO Path and Step Planning

Having defined a goal where the robot should navigate to, a plan is generated with the move_base ROS package[4]. First, a global planner based on the Dijkstra algorithm is employed to search for an optimal, obstacle free trajectory. This trajectory is fed to a local planner, in our case the Timed-Elastic-Band (TEB) planner (Rösmann et al., 2013) to compute the motion-parameters which are necessary for the robot to follow the prescribed trajectory. This local planner directly considers obstacles that can unexpectedly appear (i.e., someone passing in front of the robot) and the robot's kinematic constraints. The obtained desired velocities are then transformed to desired footstep locations with our custom ROS humanoid robot step planner.

5.1.3. NAO Walk Engine

Subsequently, the desired step locations are fed to the walking engine that computes in real-time the walking pattern (Piperakis et al., 2014) and tracks that pattern using onboard proprioceptive sensing such as the IMU, joint encoder, and pressure measurements (Piperakis and Trahanias, 2016; Piperakis et al., 2018) and/or external odometry measurements (Piperakis et al., 2019a,b) along with the current contact status (Piperakis et al., 2019c), to achieve fast and dynamically stable locomotion. The latter is vital to the success of the task since the humanoid carries a significant weight (mounted LIDAR and bowl with items) and still manages stable omnidirectional walk. The same module is also responsible for maintaining NAO's balance during fruit filling.

5.1.4. Jaco Motion Planning

For the Jaco arm, safe and accurate pick and place actions for the end-effector are learned through an offline imitation process as proposed in Koskinopoulou and Trahanias (2016) and Koskinopoulou et al. (2016). Those actions are executed via inverse kinematics in order to pick all requested breakfast items and place them in the bowl carried by the NAO robot.

5.1.5. Jaco Object Detection

The actions are triggered by visual detection of the corresponding items with an RGBD camera based on their color information with the cmvision_3d ROS package[5]. First, a detected utensil is picked by JACO and afterwards is placed when the bowl is detected.

5.2. Experimental Setup

To examine the performance of the system in the real-world, twenty volunteer FORTH employees have been recruited to interact with the robots, following the scenario summarized in section 3. In particular, the cohort for the sHRI study consisted of 14 men and 6 women with an average age of 34.5 ± 4.6 years (range, 27–45 years).

Significant variations have been observed in the times spend by the users to implement the table cleaning task. In this context, the time spent by a user is directly correlated to his/her efficiency for the task (see Equation 4) estimated for each participant. **Table 1** summarizes task completion times per participating agent for each run with a different user. Clearly, in all cases the NAO-JACO pair has accomplished to successfully and timely deliver the fruits menu to the human. In most cases the robots complete their tasks prior to the human, as evidenced by the comparison of the last and third-to-last columns.

To classify the experiment participants based on their performance, we use a k-means clustering approach to identify participants with similar behavioral characteristics. In particular, for each participant, the table cleaning time and their average efficiency are provided as input to the k-means algorithm. Multiple clustering arrangements have been explored, assuming two, three, four, and five clusters. The three-clusters arrangement is observed to group human behavioral features with sufficiently low classification cost (see **Figure 6A**). The efficiency distribution of each group is additionally depicted in **Figure 6B**, along with the relevant means.

The three clusters are assigned the labels Inattentive, Normal, and Dedicated as an implicit but representative description of the

---

[2] ROS OpenSlam Gmapping http://wiki.ros.org/gmapping

[3] ROS Adaptive Monte Carlo localization, https://github.com/ros-planning/navigation

[4] ROS move_base http://wiki.ros.org/move_base

[5] ROS cmvision_3d https://github.com/OSUrobotics/cmvision_3d

TABLE 1 | Real-User experiment data.

	NAO to JACO		Place 1st fruit		Place 2nd fruit		Place 3rd fruit		NAO to human		Table cleaning	
User	**Start**	**End**	**Start**	**End**	**Start**	**End**	**Start**	**End**	**Start**	**End**	**Start**	**End**
U1	1	55	1	66	67	108	109	150	152	190	1	185
U2	1	52	1	64	65	100	101	139	141	183	1	182
U3	1	52	1	65	66	105	106	141	143	182	1	191
U4	1	51	1	63	64	101	102	139	141	181	1	188
U5	1	57	1	74	75	111	112	152	154	192	1	194
U6	1	54	1	71	72	107	108	153	155	190	1	178
U7	1	55	1	66	67	105	–	–	107	145	1	155
U8	1	52	1	64	65	109	–	–	110	152	1	146
U9	1	48	1	59	60	104	–	–	105	142	1	168
U10	1	49	1	64	65	102	–	–	104	144	1	171
U11	1	51	1	67	68	107	–	–	108	149	1	147
U12	1	55	1	68	69	108	–	–	110	150	1	163
U13	1	47	1	61	62	101	–	–	103	142	1	161
U15	2	53	2	68	69	112	–	–	113	153	1	157
U15	1	53	1	65	66	110	–	–	111	155	1	152
U16	1	55	1	65	–	–	–	–	67	115	1	121
U17	2	54	2	62	–	–	–	–	65	117	1	134
U18	1	51	1	61	–	–	–	–	63	112	1	129
U19	1	53	1	64	–	–	–	–	67	118	1	123
U20	1	48	1	60	–	–	–	–	63	110	1	127

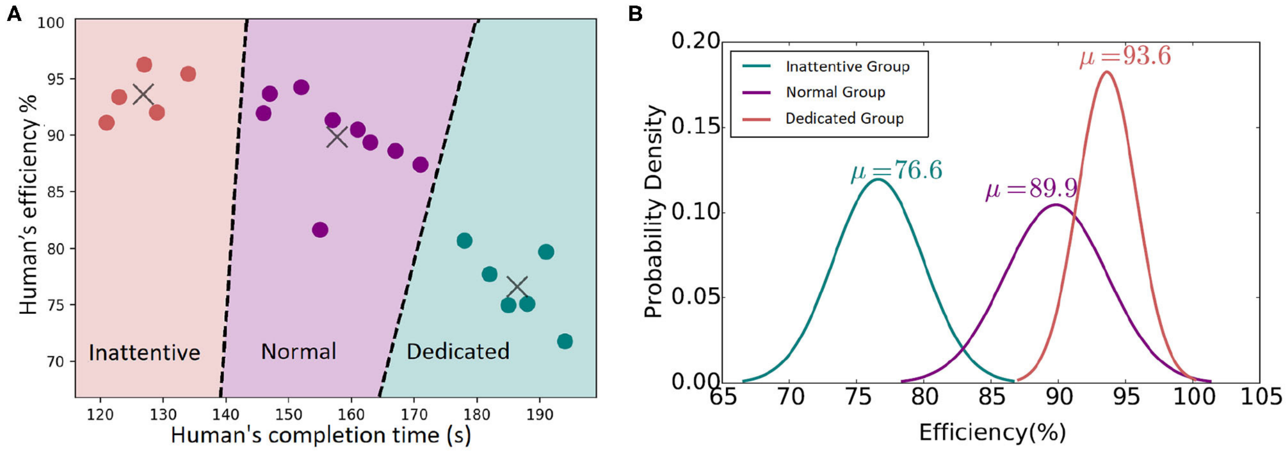


FIGURE 6 | (A) Human's behavior classification. **(B)** Normal distributions of each group's efficiency.

different human behavioral profiles observed. The grouping of human participants into Inattentive, Normal, and Dedicated is reflected in **Table 1** presentation and has been further adopted in the present work as a means to provide a more informative analysis on the evaluation of the system in human-robot interaction as presented below in section 5.3.

In addition to real-user experimentation and in order to explore the performance of the composite system in a broad range of situations and user profiles, a simulation environment has been implemented, which facilitates rigorous quantitative assessment of the proposed time-aware sHRI approach. To adequately assess the flexibility of the proposed solution, we simulate human agents assuming the same three types of user profiles, namely, Inattentive, Normal, and Dedicated, as they have been observed in the real-world experiments. The details of the underlying experimental procedure are given in section 5.4.

5.3. Real User Evaluation

5.3.1. Memory-Based Inference

The scenario assumes the inference of the breakfast preferences of the human, based on past experiences. To this end, the system capitalizes on the probabilistic inference capacity of the episodic

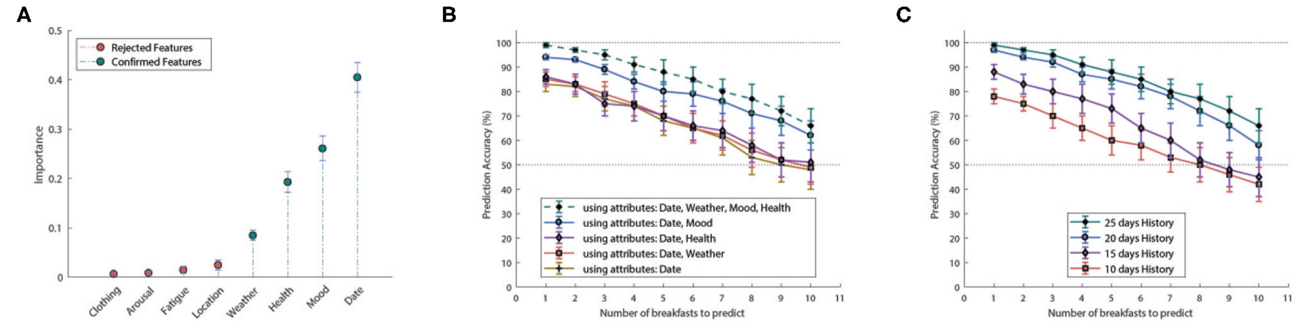


FIGURE 7 | (A) Feature importance according to Boruta feature selection algorithm. Features with red color are characterized as non-important and are rejected, whereas features with green color are characterized as important and are accepted. Inference accuracy w.r.t. prediction days and **(B)** number of attributes, **(C)** training period.

memory module to predict the user's breakfast choice, after considering the menu combinations he/she had in the past. The foreseen breakfast menu is fed to the planner which guides the two robots in fetching breakfast items and delivering them to the collaborating human, at the right time.

A data collection procedure has been adopted to provide the ground truth for assessing the performance of the episodic memory inferencing. In particular, we asked the 20 participants to provide their breakfast preferences, i.e., a selection of three fruits among six available fruit options (20 possible triplets), for 35 consecutive days. Additional information was also provided, i.e., current date, weather conditions, scenario location, user's clothing, state of arousal, fatigue, health, and mood, summing up to 8 attributes. We divided the dataset into two parts: the first 25 days are stored in memory as past experiences in the form of multi-graph episodes, while the last 10 serve as the test set for system predictions.

The data considered as "past experiences" are used for training the HMM inference engine. Specifically, in order to infer the user preferences for the *ith* day ($i > 25$) the memory is queried to obtain insight on the relevant breakfast menus the user had in the previous $1, \ldots, i-1$ days. This information is used to train an HMM, which is employed to predict user's breakfast choices on the *ith* day. The actual user choice at the given day is used as ground truth for assessing the success of breakfast predictions.

The inferencing mechanism has been evaluated against multiple configuration setups. In particular, we assessed the effectiveness of the HMM-based inferencing by making predictions of the users' breakfast menu preferences for a period of 1–10 days ahead (i.e., days 26–35). For performance enhancement we used only those attributes which are statistically significant to the system. In particular, **Figure 7A** illustrates the importance of the observed features, as computed by the Boruta feature selection algorithm. Important features, denoted with green color, are accepted to be used for state inferencing, whereas non-important features, illustrated with red color, are rejected.

In order to assess the performance of the inference mechanism, we conducted two sets of experiments with (i) varying input configurations of the four important attributes, namely date, weather conditions and user's mood and health, and with (ii) varying HMM training configurations using the most recent 10, 15, 20, or 25 days of training "history" (**Figures 7B,C**, respectively). In the first case, the HMM has been trained using 25 days of "history" and the corresponding combination of attributes, while in the second case the whole set of important attributes has been used for training, along with the corresponding "history."

Regarding the first set of experiments, **Figure 7B** demonstrates how the different attributes, i.e., the HMM observations, affect system performance. As observed, not all attributes have the same effect on the accuracy of the inference mechanism, as also implied by the relevant importance. For example "Weather Conditions" and "User's Health" play a minimal role, in contrast to the "User's Mood" which significantly improved inference performance. On the other hand, **Figure 7C** highlights the impact of the training period, i.e., number of past days ("history") used for the HMM training, on the inference accuracy. Clearly, performance is improved as the number of past days included in the training increases, i.e., more information is provided to the HMM.

In short, the proposed inference mechanism, has made highly accurate breakfast menu predictions, compatible with the personalized preferences of the individual users. Naturally, prediction accuracy decays as the looking ahead period, i.e., the period for which the mechanism is required to make predictions, extends to the future. Nevertheless, inference accuracy remained above 90% for the first 4 days, while managing to provide with adequately accurate prediction (above 80%) for a period of up to 7 days.

5.3.2. Duration Prediction

GTMs are used to predict the duration for the table cleaning task implemented by a human. For this reason, we deploy a GTM that is able to estimate accurately the time required for a human to finish the task. The experimental setup consists of a room, containing a table. A logitech HD camera is mounted at 2 m above the floor, in order to have visibility of the whole surface.

To detect the progress for the wiping the table activity, we determine a table region that designates the area to be wiped. In addition, the task observation module employs the output of the tracking module (i.e., location of the sponge), in order to identify what percent of the extracted plane has been wiped. To accomplish this, we use a hitmap as a matrix with dimensions equal to the table's width and length. As the sponge wipes the table, the matrix cells, whose rows and columns correspond to the $\{x, y\}$ coordinates of the sponge's position, are being updated from 0 to b (a scalar value), in order to indicate that the surface in these coordinates has been wiped. The value b ranges from [0 to 10] indicating the strength of the sponge while cleaning. In our experiments, b is set to 6 to match the used sponge strength. The sum of the matrix cells, divided by the product of the matrix's rows, columns and value b, provides the percentage of the surface being wiped. The value b is used to reflect the fact that when wiping a surface, one usually wipes the same area more than once. Therefore, while wiping -and updating the table matrix- one should only consider a region of the table clean if it has been wiped over b times.

To analyze the human activity we observe the wiping motions by tracking the center of the sponge, using the color based tracking framework proposed in Henriques et al. (2012). We then calculate the motion vector changes in each wiping segment, and use them to identify new primitive movements. These primitives are labeled according to the effect they have on the task progress, and used as predictors for the activity. Hence, having obtained the task progress, GTMs employ information from the observed primitives to detect the intervals that correspond to each primitive. Frequently occurring intervals within the activity, are used as predictors for the task progress. In **Table 2**, we illustrate the averaged results obtained from the three different user groups.

TABLE 2 | Average duration estimates, ground truth duration, and error measured during the wipe the table experiment, for the Inattentive, Normal and Dedicated user profiles.

User profile	Av. Predicted time	Av. Gr. truth	Error [sec]
Inattentive	177.3	186.2	8.9
Normal	149.2	143.8	5.4
Dedicated	136.9	128.6	8.3

As can be verified by the obtained results, duration predictions are robust since they fall below 10% of the overall activity duration for all three user groups. Therefore, they can support the implemented scenarios.

5.3.3. Evaluation of DP-Driven sHRI

The current section focuses on the evaluation of the Daisy Planner module, used to coordinate the activities of the agents involved in the timely breakfast delivery scenario. The behavior to be implemented by the three agents is separated into five tasks represented by five distinct petals on a Daisy Plan, as shown in **Figure 8**. The tasks are further split into actions as tabulated in **Table 3**. For all three agents, the same table shows the (min, max) times of action execution -as previously mentioned in section 4.3—and the corresponding efficacy level represented by the numbers 1 (lowest), 3, 5, 7, 9 (highest). Efficacy values are defined by the experimenter, prior to the actual experimentation. The manual setup of the planner rises some scalability issues when addressing incrementally more complex collaborative problems, since the skills of the individual agents and how they fit to the domain tasks have to be explicitly defined. Still, this is largely unavoidable and to the best of our knowledge, there is no multi-agent collaboration method that assumes minor input from the experimenter. On the other side, the current approach relies on common knowledge about the application and the separation of the composite behavior into tasks. Thus, the DP setup can be rather straightforwardly implemented since it does not assume sophisticated or difficult to obtain prior knowledge. For example, it has been very easy to employ DP for the coordination of two similar (Maniadakis et al., 2016b), or heterogeneous robots (Maniadakis et al., 2016a) in different application domains.

As discussed above, 20 different users have been involved in the real-world experimentation with the JACO and NAO robots,

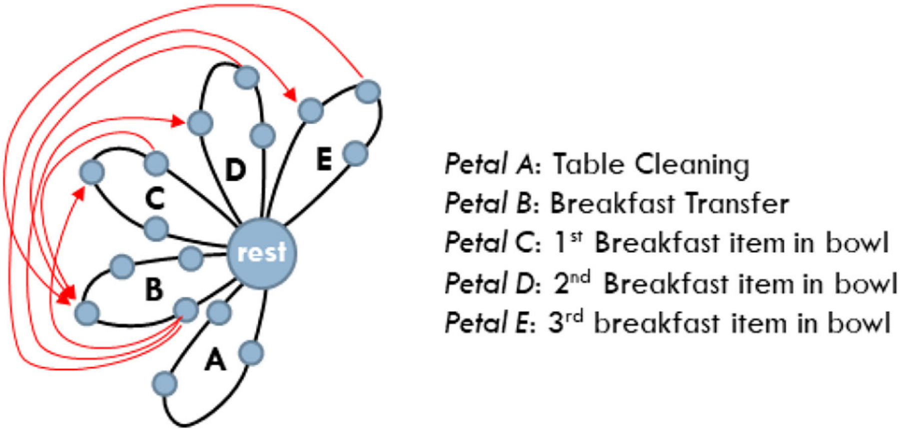


FIGURE 8 | The daisy representation of the tasks involved in the timely breakfast delivery scenario. Constraints among actions are depicted in red.

TABLE 3 | Agents' time and error level for each action.

Action	**Action**	**Human**		**NAO**		**JACO**	
Code	**Name**	**Time**	**Efficacy**	**Time**	**Efficacy**	**Time**	**Efficacy**
A1	Message: "Clean the table"	[1,2]	9	[1,2]	9	[1,2]	9
A2	Wipe the table	[122,197]	9	NaN	1	NaN	1
A3	Message: "Thank you"	[1,2]	9	[1,2]	9	[1,2]	9
A4	-Rest	[1,2]	9	[1,2]	9	[2,3]	9
B1	Move to fruit shelf	[3,5]	9	[42,57]	7	NaN	1
B2	Wait bowl filling	[14,93]	9	[14,93]	7	[14,93]	9
B3	Move to the table	[3,6]	9	[33,48]	7	NaN	1
B4	Deliver breakfast	[1,3]	9	[13,25]	7	[11,15]	9
B5	-Rest	[1,2]	9	[1,2]	9	[2,3]	9
C1	Grasp fruit1	[1,3]	9	[38,56]	3	[11,15]	7
C2	Pick fruit1	[1,2]	9	[11,15]	5	[6,8]	9
C3	Place fruit1 in bowl	[2,4]	9	[42,73]	3	[12,19]	9
C4	-Rest	[1,2]	9	[1,2]	9	[2,3]	9
D1	Grasp fruit2	[1,3]	9	[38,56]	3	[11,15]	7
D2	Pick fruit2	[1,2]	9	[11,15]	5	[6,8]	9
D3	Place fruit2 in bowl	[2,4]	9	[42,73]	3	[12,19]	9
D4	-Rest	[1,2]	9	[1,2]	9	[2,3]	9
E1	Grasp fruit3	[1,3]	9	[38,56]	3	[11,15]	7
E2	Pick fruit3	[1,2]	9	[11,15]	5	[6,8]	9
E3	Place fruit3 in bowl	[2,4]	9	[42,73]	3	[12,19]	9
E4	-Rest	[1,2]	9	[1,2]	9	[2,3]	9

in order to assess the capacity of the composite time-informed system to support sHRI. For each user, DP is fed with the three breakfast items he will most likely be interested in at the given day, as they are predicted by the episodic memory inferencing. Accordingly, in the actual DP plan, fruit-1, fruit-2, and fruit-3 are substituted by the actual fruits to be served, i.e., kiwi, orange, banana, and so on. Moreover, the GTM-based estimate of human completion time and human efficiency on the table cleaning task is used as real-time input into the planner in order to actively adapt synchronization of the three agents.

Graphical representations of task and action execution in different experimental sessions—one for each human profile—are depicted in **Figure 9**. In particular, **Figure 9A** summarizes interaction with user 4 of **Table 1**, who exhibits inattentive performance. As shown in the figure, the JACO robotic arm grasps and holds the first fruit (actions C1, C2), waiting the humanoid to arrive in a reachable area. As soon as the humanoid stops navigation (action B1), it waits for fruit filling, taking care of balancing issues (action B2). The robotic arm places the fruit it already holds in the bowl (action C3) and rests (action C4) waiting further instructions by the planner. The slow performance of the inattentive user provides enough time for JACO to add two more fruits in the humanoid's bowl (actions D1–D4 and E1–E4). As soon as all fruits are placed in the transfer bowl, NAO navigates toward the human (action B3), to deliver breakfast fruits (action B4) and rest (action B5). Almost the same time, the human completes cleaning (action A2), he gets fruits (action A3), and he is ready to enjoy breakfast (action A4).

Human-robot interaction in the case of user 11 who exhibits a normal performance profile is summarized in **Figure 9B**. The procedure followed is similar to the one summarized above, but now there is time for two fruits to be added in the bowl (actions C1–C4, and D1–D4). The humanoid robot delivers the fruits with a short delay of 2 s. The unfolding of multi-agent collaboration for the given user can be visualized in high resolution at https://youtu.be/1v4r0Xj8SF8.

Finally, **Figure 9C** summarizes the case of user 16 of **Table 1**, with an Dedicated performance profile. The high efficiency of this particular user is identified by the GTM, which informs accordingly the DP. The latter foresees that there is enough time for only one fruit to be added in the humanoid's bowl by the robotic arm (actions C1–C4). Immediately after that and in order to avoid delivery delay, the humanoid is instructed to navigate to human (action B3). The fruits are delivered on time and the user is ready to enjoy breakfast.

5.3.4. User Satisfaction

To obtain insight on the users' view on the experiment we used a post-trial questionnaire consisting of five Likert statements assessed in the scale "strongly disagree," "weakly disagree," "neutral," "weakly agree," and "strongly agree." More specifically, the following five Likert statements are examined to reveal user satisfaction:

Q1—User satisfaction: "The robots have delivered the right breakfast."

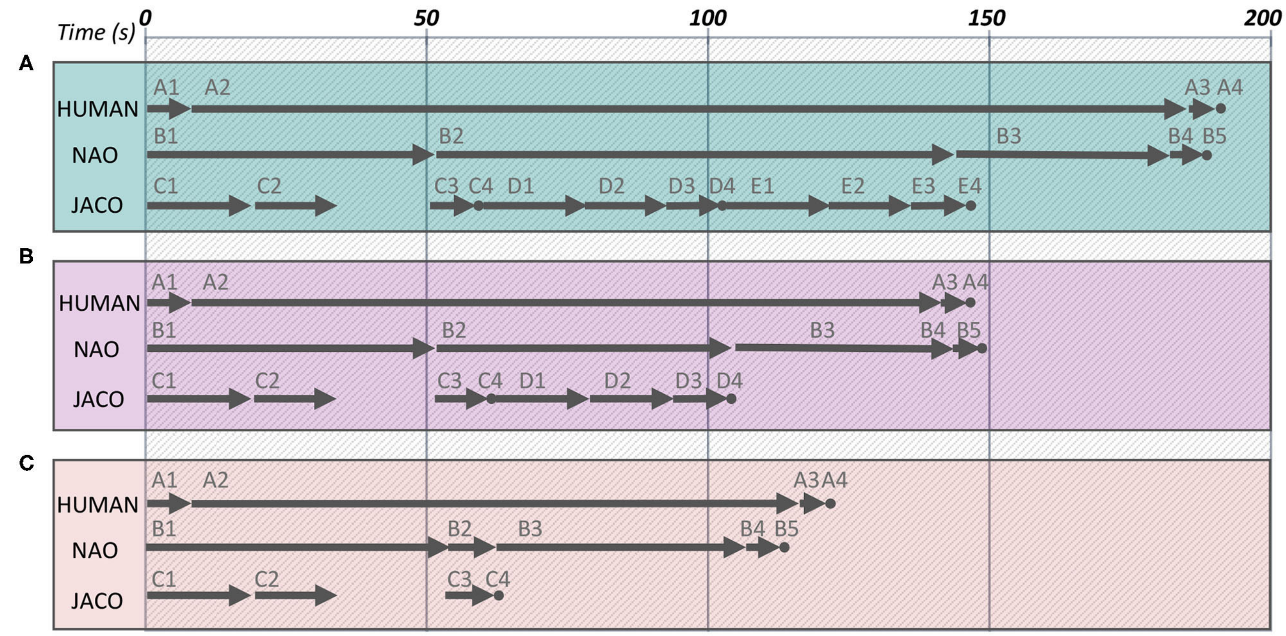


FIGURE 9 | Indicative time distribution of individual tasks for breakfast serving. **(A)** Inattentive group, **(B)** Normal group, and **(C)** Dedicated group. Rest actions, of approximately 1 s, are marked with dots.

TABLE 4 | Overview of the responses provided by participants in Q1–Q5.

Response	Q1	Q2	Q3	Q4	Q5
Strongly disagree	0	0	0	0	0
Weakly disagree	0	0	2	0	0
Neutral	3	4	5	3	2
Weakly agree	9	6	9	11	4
Strongly agree	8	10	4	6	14

Q2—User satisfaction: "The robots have delivered breakfast at the right time."
Q3—Human oriented performance: "The implementation of robot tasks was adapted to my performance."
Q4—Performance expectancy: "Robots performed better than expected."
Q5—Attitude toward using technology: "Time informed interaction is crucial for domestic robot applications."

Immediately after the experiment participants are informed that the current study is focused on time-aware multi-agent interaction. Then, they are asked to provide their individual opinions on the success of the system by filling out the aforementioned questionnaire. The aggregated results of participant responses on each question are summarized in **Table 4**. Clearly the participants expressed a positive opinion on the success of the experiment and the performance of the composite system.

A one-way MANOVA revealed a significant multivariate main effect for the type of participants, Wilks' $\lambda = 0.354$, $F_{(4,15)} = 6.82$, $p < 0.002$. This has been further confirmed by examining MANOVA nova separately for each participant type. In particular, statistically significant effects have been revealed for Inattentive Wilks'$\lambda = 0.253$, $F_{(4,15)} = 11.063$, $p < 0.0002$, for Normal Wilks' $\lambda = 0.233$, $F\ _{(}4, 15) = 12.315$, $p < 0.0001$ and for Dedicated Wilks' $\lambda = 0.356$, $F_{(4,15)} = 6.756$, $p < 0.002$. The above indicate that the answers provided by participants to Q1–Q5 are affected by their own performance on the task.

The comparative study of the answers' means revealed significant statistical difference in the satisfaction of the Dedicated and Inattentive groups, while none of them was significantly different in comparison to the Normal group. In particular, the comparison showed that the system makes the Dedicated users more satisfied than the Inattentive ones, which is due to the current parameterization of the system targeting the minimization of robot delivery delay. The latter, i.e., the robots' task execution speed, is the factor which mostly affects the overall user experience, and is yet another strong indicator of the significance of "time" in sHRI sessions.

5.4. Quantitative Assessment

To proceed in the detailed quantitative assessment of the system, we have implemented a simulation environment which assumes simulated humans to interact with the two robots. We explore system performance in a broad range of situations and user profiles, by assuming three sets of 200 simulated users, each in accordance to the Inattentive (L), Normal (N), and Dedicated (A) profiles. In particular, to obtain GTM functionality within the simulated environment we use the recorded table cleaning data from real-users as they are classified in the three user profiles. To develop more "simulated" users, each data set is scaled by α%, with α randomly specified in the range $[-10, 10]$. Thus, for each simulation run, a real human data set is randomly selected from the corresponding group L, N, or A and scaled to represent the simulated human behavior.

In close association to the unfolding of the scenario in the real world, the experiment assumes the GTM to provide DP with estimates of the hypothetical human table-cleaning duration. Subsequently, the planner attributes tasks to the agents involved in the scenario. The agents implement their tasks as indicated by the user profile and the parameters of the given run. In order to satisfy temporal constraints, the generated plans are adapted online incorporating the latest GTM estimates of the human cleaning behavior. The scenario completes with the delivery of the breakfast menu to the simulated human.

The completion times for the table cleaning and breakfast delivery tasks (by the human and the NAO robot, respectively) for the 200 simulated runs are depicted in **Figure 10A**. The three plots correspond to the L, N, A user profiles assumed in the simulation runs. It is noted that NAO's duration is strongly biased by the number of fruit items to be served to the human. More specifically, each time JACO grasps a fruit to be added in NAO's bowl it needs approximately 35 s. This explains the quantization of NAO's completion time in approximately 188 s when delivering three fruits, 146 s when delivering two fruits, and 113 s when delivering only one fruit. According to the results, in the case of the inattentive participant three items are usually served (92% three items, 8% two items). In the case of the normal human behavior two fruits are typically delivered (83% two items, 17% one item). Finally, in the case of dedicated participants only one item is commonly served (98% one item, 2% two items). This is because of the strict time constraints and the need to complete fruit delivery prior to the completion of the table cleaning task by the human.

Overall, according to **Figure 10A**, in the case of normal and inattentive users, both humans and robots have similar completion times. In 65.0% of the runs in the case of inattentive user the robot is slightly delayed in comparison to the human. As later explained (see **Figure 10B**) the relevant delays are very short. The same is also true in 72.5% of the runs in the case of normal human behavior. This is because the robots give priority on fetching more fruits undertaking the risk of a very short delay in the delivery of the breakfast. Turning to the case of the dedicated human profile, the robots generally complete their task much earlier than the human, as assumed by the rather strict request to deliver breakfast prior to table cleaning. In only 4 out of the 200 runs the robot is delayed in comparison to the human. A close look on the unfolding of the 4 mentioned runs shows that delays are introduced due to the occasionally very fast navigation of NAO close to JACO which makes the planner believe there is enough time for placing two breakfast items in NAO's bowl, which actually proves that is not the case.

Figure 10 summarizes the behaviors described above. As indicated by the relevant Gaussian distributions the implement environment.

5.4.1. Performance Metrics

To obtain insight on system's performance in association to the objective of fluent sHRI, the obtained results are quantitatively assessed using three metrics namely inter-module synchronicity, HRI synchronicity, and human idle time.

A. Inter-module Synchronicity. A key measure for assessing the effectiveness of the proposed framework regards its ability to exploit the time available for robot action, as it is predicted by the GTM. In the examined scenario this regards spending the predicted available time for fetching and transferring the maximum number of fruits to the human.

We introduce the inter-module synchronicity metric $e_{im-sync}$ which aims to reveal the temporal coupling of system modules by contrasting the expected time of tasks implementation to the actual time spent. The metric is defined as follows:

$$e^x_{im-sync} = 1 - \mathbb{E}\left\{ \frac{|\boldsymbol{t}^x_{H,p} - \max\left(\boldsymbol{t}^x_H, \boldsymbol{t}^x_N\right)|}{\max\left(\boldsymbol{t}^x_H, \boldsymbol{t}^x_N\right)} \right\} \tag{5}$$

where $x \in \{L, N, A\}$, $\boldsymbol{t}^x_{H,p}$ is the early prediction of human completion time and $\boldsymbol{t}^x_N$, $\boldsymbol{t}^x_H \in \mathbb{R}^{200}$ are the actual completion times of NAO robot and human agent, respectively, for the total of 200 simulation runs.

The obtained results are summarized in **Table 5** (Cols. 2-3). The high accuracy and low uncertainty values observed for all three human profiles are explained by the largely accurate predictions by the GTM module and the inherent flexibility of the collaboration plans developed by DP, which are sufficiently adapted to the actual implementation and the temporal characteristics of multi-agent synergistic performance.

B. HRI Synchronicity. The current metric focuses on the synchronicity of the composite artificial system with the real world. In particular, the metric describes how well human and robotic activities are synchronized, that is what is the average time that one side has to wait for the other. Synchronicity, $\boldsymbol{e}_{hri-sync}$, is defined as follows:

$$\boldsymbol{e}^x_{\text{hri}-\text{sync}} = \boldsymbol{t}^x_N - \boldsymbol{t}^x_H \tag{6}$$

where symbols are as above. As witnessed by the HRI synchronicity values shown in **Table 5** (Cols.4-5), the robotic agents are effectively synchronized with the ongoing procedures of the external environment. According to the same table, the less accurate synchronization is observed in the case of the dedicated user profile, which clearly indicates that, for the given profile, the system prefers to complete earlier the robotic task in order to minimize the risk of a possible human waiting. For the other two profiles, the robotic agents are quite accurately synchronized to humans.

The relevant observations are also reflected to the particularly low idle time of dedicated humans, which increases in the case of the Normal and Inattentive profiles (second plot of the figure).

C. Human Idle Time. To enhance human experience in sHRI sessions, the implemented system should ideally minimize the human waiting time and thus improve the responsiveness of the composite system to human requests. Similar to Hoffman (2013), we use the *Idle Time* metric to assess system performance. In particular, the human idle time $\boldsymbol{e}^x_{\text{h}-\text{idle}}$ is defined as follows:

$$\boldsymbol{e}^x_{\text{h}-\text{idle}} = \max\left(\boldsymbol{t}^x_H, \boldsymbol{t}^x_N\right) - \boldsymbol{t}^x_H \tag{7}$$

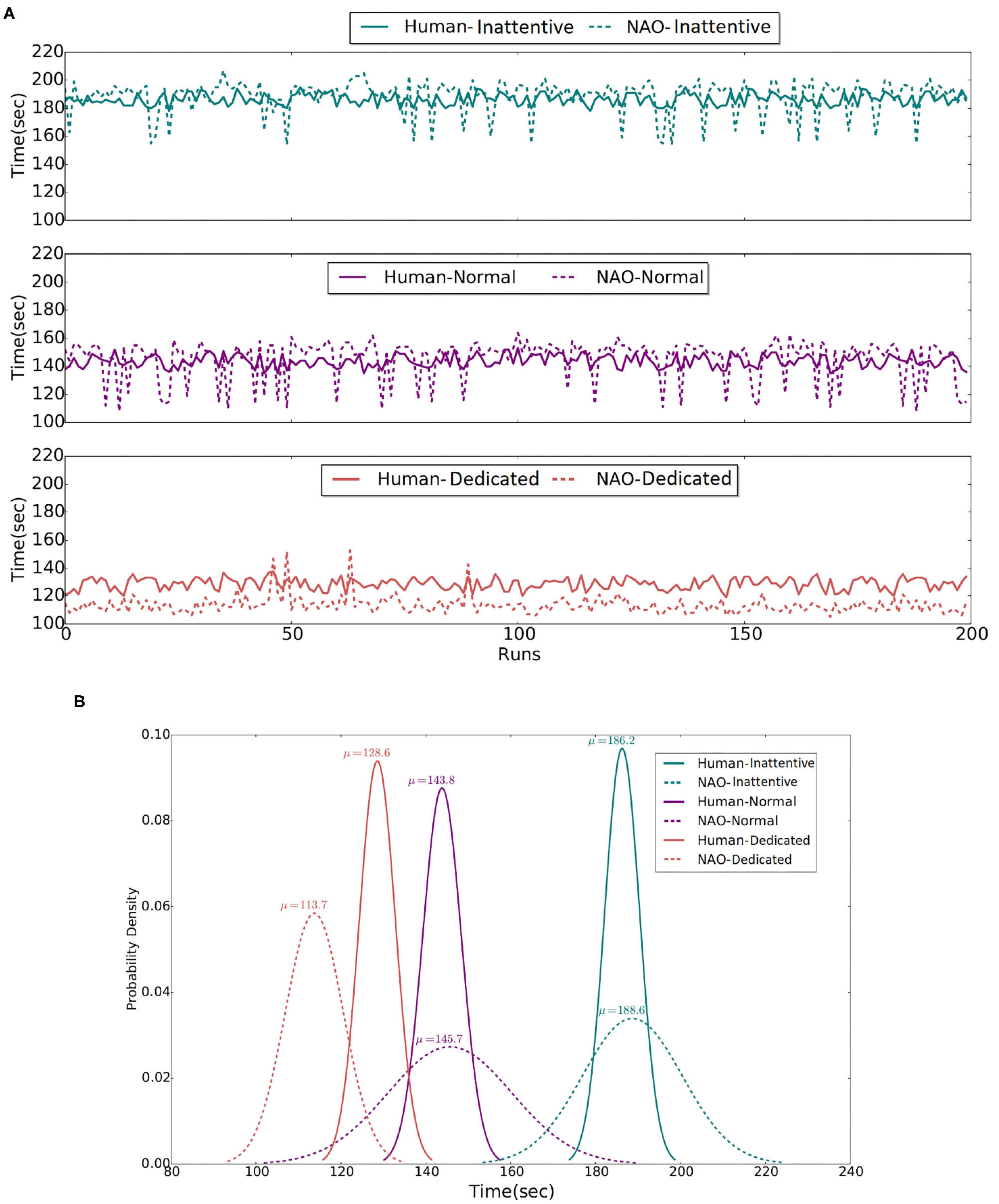


FIGURE 10 | (A) Actual times for Human-agent (lines) and NAO robot (dotted-lines) over 200 runs for the three groups. **(B)** Gaussian distributions of actual times for Human-agent (lines) and NAO robot (dotted-lines) over 200 runs for the three groups. Green, purple, and red lines indicate the Inattentive, Normal, and Dedicated group, respectively.

TABLE 5 | Performance metrics evaluation.

	Inter-module synchronicity		HRI synchronicity		Human idle time	
User profile	$\mu(e_{im-sync})(\%)$	$\sigma^2(e_{im-sync})(\%)$	$\mu(e_{hri-sync})$	$\sigma^2(e_{hri-sync})$	$\mu(e_{h-idle})$	$\sigma(e_{h-idle})$
Inattentive	98.3	0.07	2.4	126.34	5.7	31.7711
Normal	96.8	0.12	−1.9	174.72	6.3	32.1652
Dedicated	94.1	0.04	−14.8	52.38	0.3	5.5752

where symbols are as above. As shown in **Table 5** (Cols.6-7) in the case of inattentive and normal user profiles, there is a short human waiting which averages to 5.7 s in the former case and 6.3 s in the latter. However, in the context of human daily activities, durations in the range of few second are typically considered very small and do not annoy humans. In the case of the dedicated user profile, the robotic tasks typically complete earlier than the human. The only 4 out of 200 cases that robots get delayed result into the average of 0.3 s human idle time, which is particularly low and satisfactory for humans.

6. CONCLUSIONS

The long-term symbiotic interaction between humans and robots has tremendous potential as the robots bond with people, and can significantly affect humans' daily activities. Despite the increasing research endeavors devoted to human-robot synergies, we still know little about building systems that function smoothly and effectively within the context of prolonged companionships.

The integration of "sense of time" into a robotic system is at the core of a fluent sHRI, since it traverses almost every aspect of the relevant interactions. In this work, we examined the role of time focusing on three major disciplines of human-robot interaction: (i) past episode storage and experience-based inferring, (ii) activity duration prediction and human efficiency estimation, (iii) multi-agent coordination for synergetic action planning. The integrated performance of the relevant modules (i) implements time-inclusive criteria to match (assign) tasks to agents, coordinating heterogeneous agents to perform as an effective team, (ii) combines user information referring to different temporal granularities by blending the long-term, memory-inferred user preferences with the short-term, real-time predicted user expectations, (iii) monitors the dynamic scenario progress in real time to support the situated adaptation of multi-agent interaction, thus being particularly useful in sHRI scenarios.

Each module (past/episodic memory, present/Generative Time Models, future/daisy planner) benefits from the use of time-informed symbolic representation of past sHRI episodes facilitates highly accurate predictions about future scenario unfoldings. Our ongoing and future work focuses on the advancement of the individual modules in the directions shortly summarized below:

- Episodic Memory. The analysis of past data may provide significant insight into the needs and preferences of the individual users. In this direction we are currently considering mechanisms to exploit real-time robot experiences that are stored in memory, through off-line (e.g., night-time) training and dynamic querying mechanisms that recall or infer information about the users to provide accurate and personalized predictions. At the same time, research endeavors aim to improve the efficiency of the episodic memory module itself, either by fully automating the lifecycle—i.e., update, merge or forget with respect to the importance factor—of each stored memory, or by integrating the capacity to represent and exploit higher level concepts, such as the knowledge or behavioral expertise of different users.
- Generative Time Model. Given that accurate estimates of the duration of human actions greatly facilitates fluency in HRI, the current work puts forward the association of the temporal primitives of actions with the task progress. A GTM is employed to analyze human performance and provide robust estimates about the temporal aspects of the observed activity. Future work will extend the framework to discrete primitives, in order to provide a holistic methodology on temporal predictions and will consider comparative productivity measurements with emphasis on fatigue detection.
- Daisy Planner. The orchestration of team members considering their individual skills and limitations has been beneficial for sHRI applications. Our future work focuses on the temporal interruption of task implementation to enable the adaption of robot behavior to urgent un-predicted circumstances. Early work in this direction has showed that to sufficiently address this issue, the DP must distinguish between tasks (petals) which, when interrupted, can be resumed from where they left off, and tasks which, when interrupted, must be carried out completely from the beginning. Another promising direction regards the implementation of a hierarchical daisy representation of tasks, thus extending the span of human-robot interaction from the range of minutes to the range of hours, including also the ability to merge tasks similar to Stock et al. (2015).

Beyond improvements on the three core modules, future work regards the exploitation of existing computational models of human time perception (Maniadakis and Trahanias, 2012, 2016b), the time-informed kinesthetic teaching of robots (Koskinopoulou et al., 2018, 2019), and how sense of time interacts with emotions, a rather unexplored direction that has the potential to significantly improve sHRI (Maniadakis et al., 2017).

Moreover, it is of particular importance to explore the usability of the system in HRI setups that involve more agents (both robots and humans). Along this line, planned experiments will mainly focus on teams with dynamic synthesis (i.e., humans may freely enter or leave the team) and more complex, natural interaction setups with non-fully scripted scenario evolution. Overall, we envision robotic systems that greatly capitalize on temporal cognition to seamlessly integrate into the heavily time-structured human societies.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by FORTH Ethics Committee (FEC). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors have equally contributed to the submitted work.

ACKNOWLEDGMENTS

The current work was partially supported by the EU FET Proactive grant (GA: 641100) TIMESTORM—Mind and Time: Investigation of the Temporal Traits of Human-Machine Convergence, the EU FET Open grant (GA: 754423) EnTiment: Industrial Exploitation and Market Uptake of a Temporal Cognition Toolbox for Commercial Robots, and the FORTH-HONDA Research Institute—JP joint research project (Ref. No. 350/23-6/30-12-2019): Agent Profiling—Robust Performance Analysis and Prediction by Observation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frobt.2020.503452/full#supplementary-material

REFERENCES

Adam, C., Johal, W., Pellier, D., Fiorino, H., and Pesty, S. (2016). "Social human-robot interaction: a new cognitive and affective interaction-oriented architecture," in *Social Robotics*, eds A. Agah, J.-J. Cabibihan, A. M. Howard, M. A. Salichs, and H. He (Cham: Springer International Publishing), 253–263. doi: 10.1007/978-3-319-47437-3_25

Alami, R., Chatila, R., Fleury, S., Ghallab, M., and Ingrand, F. (1998). An architecture for autonomy. *Int. J. Robot. Res*. 17, 315–337. doi: 10.1177/027836499801700402

Altmann, E., and Gray, W. (2000). "Managing attention by preparing to forget," *in Proc. Human Factors Ergon Soc. Annual Meeting* 44. doi: 10.1037/0033-295X.115.3.602

Antonelli, G., Chiaverini, S., Sarkar, N., and West, M. (2001). Adaptive control of an autonomous underwater vehicle: experimental results on Odin. *IEEE Trans. Control Syst. Technol*. 9, 756–765. doi: 10.1109/87.944470

Bakker, B., Zhumatiy, V., Gruener, G., and Schmidhuber, J. (2003). "A robot that reinforcement-learns to identify and memorize important previous observations," in *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453),* Vol. 1 (Las Vegas, NV), 430–435. doi: 10.1109/IROS.2003.1250667

Baraglia, J., Cakmak, M., Nagai, Y., Rao, R., and Asada, M. (2016). "Initiative in robot assistance during collaborative task execution," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Cham), 67–74. doi: 10.1109/HRI.2016.7451735

Barreiro, J, Boyce, M., Do, M., Frank, J., Kichkaylo, M. I. T., Morris, P., et al. (2012). "Europa: a platform for ai planning, scheduling, constraint programming, and optimization," in *4th International Competition on Knowledge Engineering for Planning and Scheduling (ICKEPS)*.

Barrouillet, P., Bernardin, S., and Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *J. Exp. Psychol*. 133:83. doi: 10.1037/0096-3445.133.1.83

Biswas, M., and Murray, J. (2015). "Robotic companionship: how forgetfulness affects long-term human-robot interaction," in *International Conference on Intelligent Robotics and Applications* (Christchurch: Springer). doi: 10.1007/978-3-319-22876-1_4

Carpanzano, E., Cesta, A., Orlandini, A., Rasconi, R., Suriano, M., Umbrico, A., et al. (2016). Design and implementation of a distributed part-routing algorithm for reconfigurable transportation systems. *Int. J. Comput. Integr. Manufact*. 29, 1317–1334. doi: 10.1080/0951192X.2015.1067911

Carpenter, G. A., and Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Comput. Vis. Graph. Image Process*. 37, 54–115. doi: 10.1016/S0734-189X(87)80014-2

Carrillo, F., and Topp, E. (2016). "Interaction and task patterns in symbiotic, mixed-initiative human-robot interaction," in *AAAI Workshop: Symbiotic Cognitive Systems* (Phoenix, AZ).

Cashmore, M., Fox, M., Larkworthy, T., Long, D., and Magazzeni, D. (2014). "AUV mission control via temporal planning," in *2014 IEEE International Conference on Robotics and Automation (ICRA)* (Hong Kong), 6535–6541. doi: 10.1109/ICRA.2014.6907823

Cashmore, M., Fox, M., Long, D., Magazzeni, D., and Ridder, B. (2018). Opportunistic planning in autonomous underwater missions. *IEEE Trans. Autom. Sci. Eng*. 15, 519–530. doi: 10.1109/TASE.2016.2636662

Cashmore, M., Fox, M., Long, D., Magazzeni, D., Ridder, B., Carrera, A., et al. (2015). "Rosplan: planning in the robot operating system," in *ICAPS*.

Castelfranchi, C. (1995). "Commitments: From individual intentions to groups and organizations," in *Proc. the International Conference on Multiagent Systems* (Jerusalem), 41–48.

Castro, V. F., Clodic, A., Alami, R., and Pacherie, E. (2019). Commitments in human-robot interaction. *arXiv*.

Cesta, A., Cortellessa, G., Fratini, S., and Oddi, A. (2009). "Developing an end-to-end planning application from a timeline representation framework," in *Proceedings of the Twenty-First Conference on Innovative Applications of Artificial Intelligence* (Boston, MA).

Chao, C., and Thomaz, A. (2016). Timed petri nets for fluent turn-taking over multimodal interaction resources in human-robot collaboration. *Int. J. Robot. Res*. 35, 1330–1353. doi: 10.1177/0278364915627291

Churamani, N., Anton, P., Brügger, M., Fließwasser, E., Hummel, T., Mayer, J., et al. (2017). "The impact of personalisation on human-robo interaction in learning scenarios," in *Proceedings of the 5th International Conference on Human Agent Interaction* (Bielefeld), 171–180. doi: 10.1145/3125739.3125756

Cimatti, A., Micheli, A., and Roveri, M. (2016). "Dynamic controllability of disjunctive temporal networks: validation and synthesis of executable strategies," in *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)* (Phoenix, AZ).

Coradeschi, S., and Loutfi, A. (2008). "Chapter 15: A review of past and future trends in perceptual anchoring," in *Tools in Artificial Intelligence*, ed P. Fritzsche (Rijeka: IntechOpen), 253–265. doi: 10.5772/6081

Das, D., Rashed, M. G., Kobayashi, Y., and Kuno, Y. (2015). Supporting human-robot interaction based on the level of visual focus of attention. *IEEE Trans. Hum. Mach. Syst*. 45, 664–675. doi: 10.1109/THMS.2015.2445856

Devin, S., and Alami, R. (2016). "An implemented theory of mind to improve human-robot shared plans execution," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction* (Christchurch), 319–326. doi: 10.1109/HRI.2016.7451768

Droit-Volet, S., and Meck, W. (2007). How emotions colour our perception of time. *Trends Cogn. Sci*. 11, 504–513. doi: 10.1016/j.tics.2007.09.008

Dubois, D., and Prade, H. (1988). *Possibility Theory. An approach to Computerized Processing of Uncertainty*. New York, NY: Plenum Press. doi: 10.1007/978-1-4684-5287-7

Fernando, S., Collins, E., Duff, A., Moore, R., Verschure, P., and Prescott, T. (2014). "Optimising robot personalities for symbiotic interaction," in *Biomimetic and Biohybrid Systems*, eds A. Duff, N. F. Lepora, A. Mura, T. J. Prescott, and P. F. M. J. Verschure (Cham: Springer International Publishing), 392–395. doi: 10.1007/978-3-319-09435-9_40

Gombolay, M., Bair, A., Huang, C., and Shah, J. (2017). Computational design of mixed-initiative human-robot teaming that considers human factors: situational awareness, workload, and workflow preferences. *Int. J. Robot. Res*. 36, 597–617. doi: 10.1177/0278364916688255

Gombolay, M., Wilcox, R., and Shah, J. (2013). Fast scheduling of multi-robot teams with temporospatial constrints. *IEEE Trans. Robot*. 34, 220–239. doi: 10.15607/RSS.2013.IX.049

Green, S. A., Billinghurst, M., Chen, X., and Chase, J. G. (2008). Human-robot collaboration: a literature review and augmented reality approach in design. *Int. J. Adv. Robot. Syst*. 5:1. doi: 10.5772/5664

Hardt, O., Nader, K., and Nadel, L. (2013). Decay happens: the role of active forgetting in memory. *Trends Cogn. Sci*. 17, 111–120. doi: 10.1016/j.tics.2013.01.001

Henriques, J. F., Caseiro, R., Martins, P., and Batista, J. (2012). "Exploiting the circulant structure of tracking-by-detection with kernels," in *European Conference on Computer Vision* (Berlin; Heidelberg: Springer), 702–715. doi: 10.1007/978-3-642-33765-9_50

Ho, W. C., Dautenhahn, K., Lim, M. Y., Vargas, P. A., Aylett, R., Enz, S., et al. (2009). "An initial memory model for virtual and robot companions supporting migration and long-term interaction," in *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication* (Toyama). doi: 10.1109/ROMAN.2009.5326204

Hoffman, G. (2013). "Evaluating fluency in human-robot collaboration," in *Robotics: Science and Systems, Workshop on Human Robot Collaboration* (Berlin), 1–8.

Hourdakis, E., and Trahanias, P. (2013). Use of the separation property to derive liquid state machines with enhanced classification performance. *Neurocomputing* 107, 40–48. doi: 10.1016/j.neucom.2012.07.032

Hourdakis, E., and Trahanias, P. (2018). "A robust method to predict temporal aspects of actions by observation," in *International Conference on Robotics Automation, ICRA*. doi: 10.1109/ICRA.2018.8461067

Hunsberger, L. (2014). "A faster algorithm for checking the dynamic controllability of simple temporal networks with uncertainty," in *Proceedings of the 6th International Conference on Agents and Artificial Intelligence (ICAART)* (Angers).

Ingrand, F., and Ghallab, M. (2017). Deliberation for autonomous robots: a survey. *Artif. Intell*. 247, 10–44. doi: 10.1016/j.artint.2014.11.003

Isaacson, S., Rice, G., and Boekoel, J. C. Jr. (2019). MAD-TN: A tool for measuring fluency in human-robot collaboration. *arXiv:1909.06675v3*. Available online at: https://www.semanticscholar.org/paper/MAD-TN%3A-A-Tool-for-Measuring-Fluency-in-Human-Robot-Isaacson-Rice/8b7a45410e5ceea0fc01aa7950dd3d87cd9b6f75

Kelley, R., Tavakkoli, A., King, C., Nicolescu, M., Nicolescu, M., and Bebis, G. (2008). "Understanding human intentions via hidden markov models in autonomous mobile robots," in *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction* (New York, NY: ACM), 367–374. doi: 10.1145/1349822.1349870

Khadar, B., Rajesh, A., Madhusudhan, R., Ramanaiah, M. V., and Karthikeyan, K. (2013). Statistical optimization for generalised fuzzy number. *Int. J. Modern Eng. Res*. 3, 647–651.

Koskinopoulou, M., Maniadakis, M., and Trahanias, P. (2018). "Learning spatio-temporal characteristics of human motions through observation," in *International Conference on Robotics in Alpe-Adria Danube Region* (Patras), 82–90. doi: 10.1007/978-3-030-00232-9_9

Koskinopoulou, M., Maniadakis, M., and Trahanias, P. (2019). Speed adaptation in learning from demonstration through latent space formulation. *Robotica* 38, 1–13. doi: 10.1017/S0263574719001449

Koskinopoulou, M., Piperakis, S., and Trahanias, P. (2016). "Learning from demonstration facilitates human-robot collaborative task execution," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Christchurch), 59–66. doi: 10.1109/HRI.2016.7451734

Koskinopoulou, M., and Trahanias, P. E. (2016). "A methodological framework for robotic reproduction of observed human actions: formulation using latent space representation," in *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)* (Cancún: IEEE Press), 565–572. doi: 10.1109/HUMANOIDS.2016.7803331

Kotseruba, I., and Tsotsos, J. (2018). 40 years of cognitive architectures: core cognitive abilities and practical applications. *Artif. Intell. Rev*. 53, 17–94. doi: 10.1007/s10462-018-9646-y

Kursa, M. B., Jankowski, A., and Rudnicki, W. R. (2010). Boruta-a system for feature selection. *Fund. Inform*. 101, 271–285. doi: 10.3233/FI-2010-288

Langley, P., Laird, J., and Rogers, S. (2009). Cognitive architectures: research issues and challenges. *Cogn. Syst. Res*. 10, 141–160. doi: 10.1016/j.cogsys.2006.07.004

Laptev, I. (2005). On space-time interest points. *Int. J. Comput. Vis*. 64, 107–123. doi: 10.1007/s11263-005-1838-7

Leconte, F., Ferland, F., and Michaud, F. (2016). Design and integration of a spatio-temporal memory with emotional influences to categorize and recall the experiences of an autonomous mobile robot. *Auton. Robots* 40, 831–848. doi: 10.1007/s10514-015-9496-2

Lemaignan, S., Warnier, M., Sisbot, E. A., Clodic, A., and Alami, R. (2017). Artificial cognition for social humanrobot interaction. *Artif. Intell*. 247, 45–69. doi: 10.1016/j.artint.2016.07.002

Lin, Z., Jiang, Z., and Davis, L. S. (2009). "Recognizing actions by shape-motion prototype trees," in *IEEE 12th International Conference on Computer Vision* (Kyoto), 444–451.

Liu, P., Glas, D. F., Kanda, T., and Ishiguro, H. (2016). Data-driven HRI: Learning social behaviors by example from human-human interaction. *IEEE Trans. Robot*. 32, 988–1008. doi: 10.1109/TRO.2016.2588880

Maniadakis, M., Aksoy, E., Asfour, T., and Trahanias, P. (2016a). "Collaboration of heterogeneous agents in time constrained tasks," in *Proc. IEEE-RAS International Conference on Humanoid Robots (Humanoids)* (Cancún). doi: 10.1109/HUMANOIDS.2016.7803314

Maniadakis, M., Droit-Volet, S., and Trahanias, P. (2017). "Emotionally modulated time perception for prioritized robot assistance," in *Proc. ACM/IEEE Int. Conference on Human-Robot Interaction (HRI)* (Vienna). doi: 10.1145/3029798.3038398

Maniadakis, M., Hourdakis, E., and Trahanias, P. (2016b). Time-informed task planning in multi-agent collaboration. *Cogn. Syst. Res*. 43, 291–300. doi: 10.1016/j.cogsys.2016.09.004

Maniadakis, M., Hourdakis, M., and Trahanias, P. (2007). "Modeling overlapping execution/observation brain pathways," in *International Joint Conference on Neural Networks, 2007* (Orlando, FL), 1255–1260. doi: 10.1109/IJCNN.2007.4371138

Maniadakis, M., and Surmann, H. (1999). "A genetic algorithm for structural and parametrical tuning of fuzzy systems," in *European Symposium on Intelligent Techniques, ESIT99* (Chania).

Maniadakis, M., and Trahanias, P. (2012). "Experiencing and processing time with neural networks," in *Proc. 4th Int. Conf. on Advanced Cogn. Tech. and App* (Stará Lesná).

Maniadakis, M., and Trahanias, P. (2015). Integrated intrinsic and dedicated representations of time: a computational study involving robotic agents. *Timing Time Percept*. 3, 246–268. doi: 10.1163/22134468-03002052

Maniadakis, M., and Trahanias, P. (2016a). "Time-informed, adaptive multi-robot synchronization," in *Proc. Simulation of Adaptive Behavior (SAB)* (Aberystwyth). doi: 10.1007/978-3-319-43488-9_21

Maniadakis, M., and Trahanias, P. (2016b). When and how-long: a unified approach for time perception. *Front. Psychol.* 7:466. doi: 10.3389/fpsyg.2016.00466

Maniadakis, M., Trahanias, P., and Tani, J. (2009). Explorations on artificial time perception. *Neural Netw.* 22, 509–517. doi: 10.1016/j.neunet.2009.06.045

Maniadakis, M., Wittmann, M., and Trahanias, P. (2011). "Time experiencing by robotic agents," in *Proceedings of 11th European Symposium on Artificial Neural Networks* (Bruges), 429–434.

Melvin, J., Keskinocak, P., Koenig, S., Tovey, C. A., and Ozkaya, B. Y. (2007). "Multi-robot routing with rewards and disjoint time windows," in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems* (San Diego, CA), 2332–2337. doi: 10.1109/IROS.2007.4399625

Michael, J., and Pacherie, E. (2015). On commitments and other uncertainty reduction tools in joint action. *J. Soc. Ontol.* 1, 89–120. doi: 10.1515/jso-2014-0021

Morris, P. (2014). "Dynamic controllability and dispatchability relationships," in *Integration of AI and OR Techniques in Constraint Programming, Lecture Notes in Computer Science*, ed H. Simonis (Cham: Springer International Publishing), 464–479. doi: 10.1007/978-3-319-07046-9_33

Morris, P., Muscettola, N., and T., V. (2001). "Dynamic control of plans with temporal uncertainty," in *International Joint Conference on AI (IJCAI)* (Washington, DC).

Nunes, E., and Gini, M. (2015). "Multi-robot auctions for allocation of tasks with temporal constraints," in *AAAI Conference on Artificial Intelligence* (Austin, TX), 2110–2116.

Nunes, E., Nanjanath, M., and Gini, M. (2012). "Auctioning robotic tasks with overlapping time windows," in *International Conference on Autonomous Agents and Multi-Agent Systems* (Valencia), 1211–1212.

Nuxoll, A. M., and Laird, J. E. (2012). Enhancing intelligent agents with episodic memory. *Cogn. Syst. Res.* 17–18, 34–48. doi: 10.1016/j.cogsys.2011.10.002

Oberauer, K., and Lewandowsky, S. (2008). Forgetting in immediate serial recall: decay, temporal distinctiveness, or interference? *Psychol. Rev.* 115, 544–576. doi: 10.1037/0033-295X.115.3.544

Oberauer, K., and Lewandowsky, S. (2011). Modeling working memory: a computational implementation of the time-based resource-sharing theory. *Psychon. Bull. Rev.* 18. doi: 10.3758/s13423-010-0020-6

Pellegrinelli, S., Orlandini, A., Pedrocchi, N., Umbrico, A., and Tolio, T. (2017). Motion planning and scheduling for human and industrial-robot collaboration. *CIRP Ann.* 66, 1–4. doi: 10.1016/j.cirp.2017.04.095

Petit, M., Fischer, T., and Demiris, Y. (2016). Lifelong augmentation of multimodal streaming autobiographical memories. *IEEE Trans. Cogn. Dev. Syst.* 8, 201–213. doi: 10.1109/TAMD.2015.2507439

Piperakis, S., Kanoulas, D., Tsagarakis, N. G., and Trahanias, P. (2019a). "Outlier-robust state estimation for humanoid robots," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Macau), 706–713. doi: 10.1109/IROS40897.2019.8968152

Piperakis, S., Koskinopoulou, M., and Trahanias, P. (2018). "Nonlinear state estimation for humanoid robot walking," in *IEEE Robotics and Automation Letters, Vol. 3* (Madrid), 3347–3354. doi: 10.1109/LRA.2018.2852788

Piperakis, S., Orfanoudakis, E., and Lagoudakis, M. (2014). "Predictive control for dynamic locomotion of real humanoid robots," in *IEEE/RSJ International Conference on Intel. Robots and Systems* (Chicago, IL), 4036–4043. doi: 10.1109/IROS.2014.6943130

Piperakis, S., Tavoularis, N., Hourdakis, E., Kanoulas, D., and Trahanias, P. (2019b). "Humanoid robot dense RGB-D SLAM for embedded devices," in *IEEE International Conference on Robotics and Automation (ICRA)*: *3rd Full-Day Workshop* "*Towards Real-World Deployment of Legged Robots*" (Macau).

Piperakis, S., Timotheatos, S., and Trahanias, P. (2019c). "Unsupervised gait phase estimation for humanoid robot walking," in *2019 International Conference on Robotics and Automation (ICRA)* (Macau), 270–276. doi: 10.1109/ICRA.2019.8793598

Piperakis, S., and Trahanias, P. (2016). "Non-linear ZMP based state estimation for humanoid robot locomotion," in *IEEE-RAS International Conference on Humanoid Robots* (Cancun), 202–209. doi: 10.1109/HUMANOIDS.2016.7803278

Py, F., Rajan, K., and McGann, C. (2010). "A systematic agent framework for situated autonomous systems," in *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: Volume 2, AAMAS '10* (Taipei: International Foundation for Autonomous Agents and Multiagent Systems), 583–590.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* 77, 257–286. doi: 10.1109/5.18626

Rajan, K., McGann, C., Py, F., and Thomas, H. (2007). "Robust mission planning using deliberative autonomy for autonomous underwater vehicles," in *Proceedings of ICRA Workshop on Robotics in Challenging and Hazardous Environments* (Rome), 21–25.

Rajan, K., and Saffiotti, A. (2017). Towards a science of integrated ai and robotics. *Artif. Intell.* 247, 1–9. doi: 10.1016/j.artint.2017.03.003

Riccio, F., Vanzo, A., Mirabella, V., Catarci, T., and Nardi, D. (2016). "Enabling symbiotic autonomy in short-term interactions: a user study," in *Social Robotics*, eds A. Agah, J.-J. Cabibihan, A. M. Howard, M. A. Salichs, and H. He (Cham: Springer International Publishing), 796–807. doi: 10.1007/978-3-319-47437-3_78

Rosenthal, S., Biswas, J., and Veloso, M. (2010). "An effective personal mobile robot agent through symbiotic human-robot interaction," in *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent System (AAMAS)* (Toronto, ON), 915–922.

Rösmann, C., Feiten, W., Wösch, T., Hoffmann, F., and Bertram, T. (2013). "Efficient trajectory optimization using a sparse model," in *2013 European Conference on Mobile Robots* (Barcelona), 138–143. doi: 10.1109/ECMR.2013.6698833

Scovanner, P., Ali, S., and Shah, M. (2007). "A 3-dimensional sift descriptor and its application to action recognition," in *Proceedings of the 15th ACM International Conference on Multimedia* (Augsburg), 357–360. doi: 10.1145/1291233.1291311

Shah, J., Stedl, J., Williams, B., and Robertson, P. (2007). "A fast incremental algorithm for maintaining dispatchability of partially controllable plans," in *Proceedings of the 18th International Conference on Automated Planning and Scheduling* (Sydney, NSW), 296–303.

Sigalas, M., Maniadakis, M., and Trahanias, P. (2017a). "Episodic memory formulation and its application in long-term HRI," in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (Lisbon), 599–606. doi: 10.1109/ROMAN.2017.8172364

Sigalas, M., Maniadakis, M., and Trahanias, P. (2017b). "Time-aware long-term episodic memory for recurring HRI," in *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (Vienna), 287–288. doi: 10.1145/3029798.3038307

Smith, S., Gallagher, A., Zimmerman, T., Barbulescu, L., and Rubinstein, Z. (2007). "Distributed management of flexible times schedules," in *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (Honolulu, HI). doi: 10.1145/1329125.1329215

Stock, S., Mansouri, M., Pecora, F., and Hertzberg, J. (2015). "Online task merging with a hierarchical hybrid task planner for mobile service robots," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Hamburg), 6459–6464. doi: 10.1109/IROS.2015.7354300

Taylor, S. E., Vineyard, C. M., Healy, M. J., Caudell, T. P., Cohen, N. J., Watson, P., et al. (2009). "Memory in silico: Building a neuromimeti episodic cognitive model," in *2009 WRI World Congress on Computer Science and Information Engineering, Vol. 5*, 733–737. doi: 10.1109/CSIE.2009.978

Tsarouchi, P., Makris, S., and Chryssolouris, G. (2016). Human-robot interaction review and challenges on task planning and programming. *Int. J. Comput. Integr. Manufactur*. 29, 916–931. doi: 10.1080/0951192X.2015.1130251

Tscherepanow, M., Kühnel, S., and Riechers, S. (2012). "Episodic clustering of data streams using a topology-learning neural network," in *Proceedings of the ECAI Workshop on Active and Incremental Learning* (Kingston, NY).

Umbrico, A., Cesta, A., Cialdea Mayer, M., and Orlandini, A. (2017). "Platinum: a new framework for planning and acting," in *AI*IA 2017 Advances in Artificial Intelligence*, eds F. Esposito, R. Basili, S. Ferilli, and F. A. Lisi, 498–512. doi: 10.1007/978-3-319-70169-1_37

Wang, H., Ullah, M. M., Klser, A., Laptev, I., and Schmid, C. (2009). "Evaluation of local spatio-temporal features for action recognition," in *BMVC 2009* (London: BMVA Press). doi: 10.5244/C.23.124

Wang, Y., and Mori, G. (2009). Human action recognition by semilatent topic models. *Trans. Pattern Anal. Mach. Intell.* 31, 1762–1774. doi: 10.1109/TPAMI.2009.43

Wilcox, R., Nikolaidis, S., and A., S. J. (2013). Optimization of temporal dynamics for adaptive human-robot interaction in assembly manufacturing. *Robotics* 441, 442–450. doi: 10.15607/RSS.2012.VIII.056

Wilcox, R., Nikolaidis, S., and Shah, J. A. (2012). "Optimization of temporal dynamics for adaptive human-robot interaction in assembly manufacturing," in *Proceedings of Robotics: Science and Systems* (Sydney, NSW).

Zakay, D. (1989). Subjective time and attentional resource allocation: an integrated model of time estimation. *Adv. Psychol*. 59, 365–397. doi: 10.1016/S0166-4115(08)61047-X

Zakay, D., and Block, R. A. (1996). The role of attention in time estimation processes. *Adv. Psychol*. 115, 143–164. doi: 10.1016/S0166-4115(96)80057-4

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

BRIEF RESEARCH REPORT
published: 24 February 2021
doi: 10.3389/fpsyg.2021.574983

Gratitude Affects Inter-Subjective Synchronicity for Cognitive Performance and Autonomic Responsiveness

Michela Balconi[1,2] *and Giulia Fronda*[1,2]*

[1]*International Research Center for Cognitive Applied Neuroscience (IrcCAN), Catholic University of the Sacred Heart, Milan, Italy,* [2]*Research Unit in Affective and Social Neuroscience, Department of Psychology, Catholic University of the Sacred Heart, Milan, Italy*

Recently, social neurosciences have been interested in the investigation of neurophysiological responses related to the experience of positive emotions, such as gratitude, during social interactions. Specifically, the aim of the present research was to investigate whether gratitude related to gift exchange could favor cooperative behavior and bond construction, by improving behavioral and autonomic responsivity. At this regard, the autonomic synchronization and behavioral performance of 16 friends coupled in dyads were recorded during a joint attentional task. Gift exchange could be occurred either at the beginning or in the middle of the task. For the recording of simultaneous autonomic activity [heart rate (HR) and skin conductance level (SCL)], a hyperscanning biofeedback paradigm was used. Intra-subjective analysis showed an increase in behavioral [accuracy (ACC)] and autonomic responses (HR and SCL) when the gift exchange took place at the beginning of the task rather than in the middle. Moreover, inter-subjective analysis revealed an increase in behavioral performance and greater autonomic synchronization of HR index. The present research, therefore, shows how gratitude and trust experienced following gift exchange can modify participants' reactions by creating a shared cognition and the adoption of joint strategies.

Keywords: gratitude, heart rate, skin conductance level, biofeedback, hyperscanning

Edited by:
Peter König,
University of Osnabrück, Germany

Reviewed by:
Dan Zhang,
Tsinghua University, China
Judith M. Burkart,
University of Zurich, Switzerland

****Correspondence:***
Giulia Fronda
giulia.fronda@unicatt.it

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *30 June 2020*
Accepted: *01 February 2021*
Published: *24 February 2021*

Citation:
Balconi M and Fronda G (2021) Gratitude Affects Inter-Subjective Synchronicity for Cognitive Performance and Autonomic Responsiveness. Front. Psychol. 12:574983. doi: 10.3389/fpsyg.2021.574983

INTRODUCTION

The act of giving or receiving something can be considered as a moment of interpersonal exchange that leads to the development of important links that strengthen social relations (Mick and Demoss, 1990; Mick and Faure, 1998). Specifically, positive emotions experienced during gift exchange, such as gratitude, turn out to be important in the construction of social ties, improving subjective well-being and prosocial behavior by strengthening interpersonal relationships (McCullough and Tsang, 2004; Penner et al., 2005; Algoe et al., 2008; Froh et al., 2008; Lambert et al., 2010). In particular, the moment of gift exchange, as an act of social interaction, involves the individuals by influencing and modeling their behaviors (Golland et al., 2015), thanks to those basic mechanisms that allow to perceive, imitate, and understand others' feelings, actions, and intentions (Niedenthal, 2007; Balconi and Bortolotti, 2012, 2013; Balconi and Canavesio, 2014). As demonstrated by previous studies, indeed, when these mirroring mechanisms happen in interacting individuals, an implicit behavioral, neural, and psychophysiological activity attunement occurs (Richardson et al., 2007; Konvalinka et al., 2010, 2011; Dumas et al., 2011; Müller and Lindenberger, 2011; Hasson et al., 2012; Giuliano et al., 2015), leading neuroscientists to consider inter-agent actors as a single complex system (Balconi and Pagani, 2015; Chung et al., 2015; Balconi and Vanutelli, 2017a). Specifically, the body synchronization that is experienced between the inter-agents may be due to the sharing of positive emotional experiences during gift exchange (Chauhan et al., 2008; Balconi and Canavesio, 2013). Emotions, indeed, improve behavioral, cognitive, and affective individuals' synchronization representing the basis of prosocial behavior (Balconi et al., 2011; Balconi and Bortolotti, 2012; Vanutelli et al., 2016, 2017; Balconi and Vanutelli, 2017a).

An innovative paradigm, called hyperscanning, has been proposed to explore the synchronization that occurs between the interacting individuals during joint actions (Vanutelli et al., 2016, 2017; Balconi and Vanutelli, 2017b; Balconi et al., 2018). According to this recent paradigm, the focus is, therefore, on the recording of individuals' neurophysiological activity during various interpersonal dynamics (Schilbach et al., 2010). As demonstrated by some studies, indeed, the neural and autonomic responses of two interacting individuals can show a strong synchronization during a significant interpersonal exchange (Levenson and Ruef, 1992). Specifically, simultaneous neural activity can be recorded through the use of different techniques, such as electroencephalography (EEG; Koike et al., 2015; Balconi and Vanutelli 2017a), and neuroimaging techniques, such as functional near-infrared spectroscopy (fNIRS), which permit to observe hemodynamic activity changes in specific brain regions (Balconi and Vanutelli, 2017a). On the other hand, peripheral activity can be recorded through the use of biofeedback (Mirgorodsky et al., 2013). In particular, as consistently shown by previous work on autonomic attunement (Balconi and Vanutelli, 2017a; Vanutelli et al., 2017, 2018), the measurement of peripheral synchrony permits to evaluate the physiological synchronization (PS), defined as the covariation of autonomous measurements between dyads or interacting groups (Butler, 2011), by analyzing their changes throughout the task.

The PS index, therefore, can be used to define the intensity of the interaction between the inter-agent individuals (Hatfield et al., 1994). Moreover, the peripheral activity, considered as a response implemented by the sympathetic nervous system (SNS) to identify "fight-or-flight" responses, can provide information on individuals' emotional state (Diamond and Otter-Henderson, 2007). It turned out to be correlated to some important social, emotional, and empathic processes (Levenson and Ruef, 1992; Adolphs, 2003) and it gives information related to individuals' synchrony (Chaspari et al., 2015). For these reasons, autonomic synchronization was observed above all in cooperative behaviors (Vanutelli et al., 2016; Balconi and Vanutelli, 2017a) in significant relational ties, such as parent-child interaction (Barrett, 2006) and in the patient-therapist relationship (Marci et al., 2007). Specifically, information on individuals' emotional activation (Diamond and Otter-Henderson, 2007; Boucsein, 2012) and the quality of social interactions (Guastello et al., 2006) were provided by electrodermal activity (EDA), or skin conductance level (SCL), that is, a representative index of skin conductivity changes. In addition to EDA, the cardiovascular activity also provides information on the individual's emotional activation. In fact, it increases mainly while experiencing highly positive or negative emotions, such as happiness, joy, fear, sadness, and anger (Levenson, 1992; Sinha et al., 1992), and varies according to emotional closeness (Konvalinka et al., 2011).

These indices, therefore, can provide information related to the emotional synchronization between two individuals involved in social exchange, as revealed in previous studies that observed a higher autonomic activity [increased arousal, SCL, and heart rate (HR)] during social interactions, such as cooperative conditions (Balconi and Bortolotti, 2012). Furthermore, it was shown that the emotional attunement between inter-agents that is experienced during cooperative behaviors leads individuals to improve their behavioral performance through the releasing of dopamine in the prefrontal brain regions. Such release produces an improvement in cognitive performance on a wide range of tasks involving the use of attention and working memory (Gray, 2004; Isen, 2009; Nadler et al., 2010).

The aim of this study was, therefore, to explore the peripheral activity synchronization between two individuals, using a hyperscanning technique based on biofeedback, during the performance of a task that involved a gift exchange (consisting of a material or experiential gift) at the beginning or in the middle of the performance of a cognitive task consisting of three blocks in which was asked to recognize, among others, a target stimulus. Specifically, we hypothesized that the performance of joint action, involving the establishment of cooperative behavior, would lead to an increase in individuals' peripheral synchronization and behavioral performance. In this regard, an attentive task was devised that required pairs of subjects to cooperate by synchronizing their responses. In particular, the main aim of the present work was to investigate whether the participants' behavioral performance and peripheral activity improved following the gift exchange. Secondly, we aimed at verifying if the specific moment of gift exchange (at the beginning or in the middle of the task) had specific effects on individuals' peripheral activity and behavioral responses. Thirdly, we hypothesized that the moment of gift exchange, involving the sharing of positive emotions, such as gratitude, could strengthen individuals' behavioral performance and autonomic synchronization. Specifically, we expected this effect to be higher in the first condition in which the gift exchange took place at the beginning of the task. Finally, we expected to find different effects in the autonomic activity of the donor compared to that of the receiver before and after the gift exchange, in terms of a higher emotional engagement and a higher arousal for the donor than receiver (Flynn and Brockner, 2003; Duclos and Barasch, 2014).

MATERIALS AND METHODS

Participants

Sixteen pairs of subjects (*N* = 32, *M* = 23.34; *SD* = 1.23) involved in a friendship relationship were recruited to carry out the experiment, using the following exclusion criteria: normal or correct-normal visual acuity and absence of neurological or psychiatric pathologies, verified by specific measurements. Two dyads were excluded due to the low quality of autonomic signal. The dyads took part in the study after signing the informed consent. The research was conducted according to the Helsinki Declaration and was approved by the local ethics committee of the Department of Psychology of the Catholic University of Milan.

Procedure

To carry out the experiment, the subjects were seated in a dark room at a 60 cm from a computer.

Specifically, the subjects were asked to carry out a computerized task that involved a gift exchange at the beginning or in the middle of the task, which had to be donated by one of the subjects of the dyad (donor) to his partner (receiver). For seven couples, the gift exchange occurred before the beginning of the first part (after block 1); for the other seven pairs, it occurred at the end of the second block. Two different procedures were used: the first involved the performance of a basic condition (block 1), the gift exchange, and the execution of the other two task blocks (blocks 2 and 3), and the second provided for the unwinding of blocks 1 and 2, the gift exchange, and the carrying out of the block 3. Therefore, at the beginning or in the middle of the task (concerning the two procedural orders), they were asked to exchange a gift, which could be a material or experiential object (objects or tickets to visit a museum or a concert). Before the execution of blocks 2 and 3, after recording a basic condition of 120 s, the participants were asked to perform a familiarization activity with the activity (block 1; **Figure 1**).

Specifically, the execution of the three blocks required the participants to cooperate during the development of a selective attention task modified by a previous computerized version (Balconi and Pagani, 2015; Balconi and Vanutelli, 2016; Vanutelli et al., 2016, 2017; Balconi et al., 2018, 2019a,b, 2020; Balconi and Fronda, 2020). The task required subjects to memorize a target stimulus (triangle or circle and green or blue) that they should have subsequently recognized among others by pressing the right or left key of the computer keyboard. Specifically, the stimulus appeared on the screen for 500 milliseconds (ms) with an inter-stimulus interval (ISI) of 300 ms and an inter-trial interval (ITI) of 5,000 ms. The task required the subjects to recognize the target stimulus by synchronizing the speed and accuracy (ACC), understood as the percentage of the correct responses concerning the recognition of the target stimulus, of their responses with

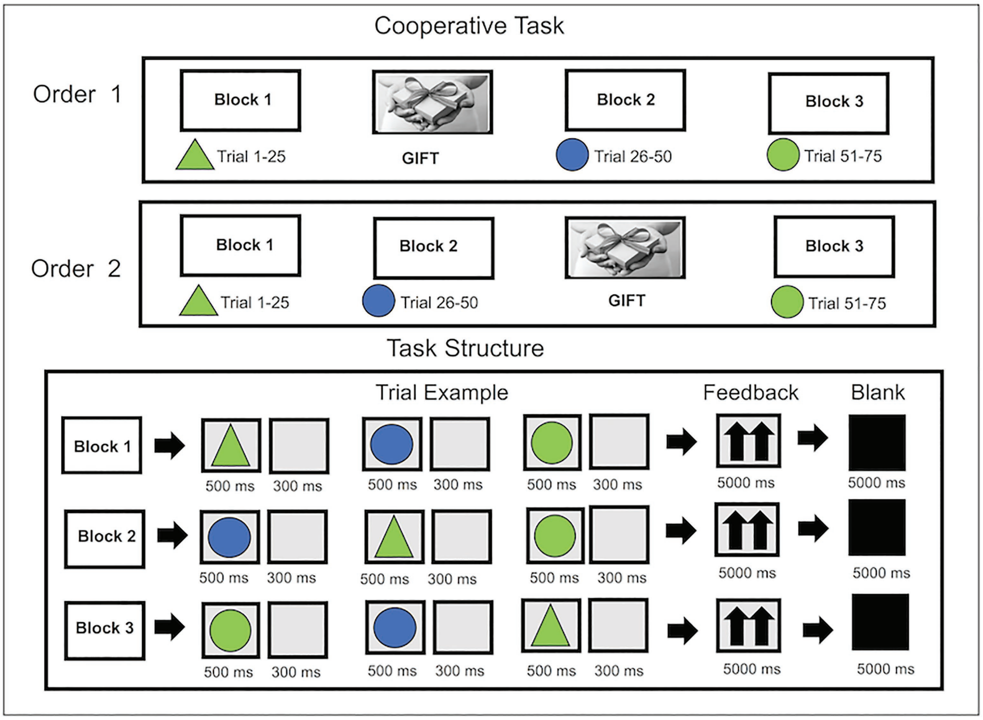


FIGURE 1 | Experimental procedure. Two different procedures were performed: order 1 composed by block 1 (a control condition), gift exchange, and blocks 2 and 3; and order 2 composed by block 1, block 2, gift exchange, and block 3. Blocks 1, 2, and 3 involve a cooperative task which consisted of a game of selective attention.

the support of some feedback indicating the individuals' cooperation level represented by two arrows pointing upward, indicating a level of cooperation over 75%. Then, the subjects were asked to temporarily associate their reaction times (RTs) to increase their perception and psychological condition of cooperation and joint task. At the end of the task, both participants were given a questionnaire in order to investigate the perception of the partner (his/her friend) and the level of couple tuning during the performance of the first and second task blocks. The questionnaire consisted of the following items: "What was the perception of your workmate in the first phase of the game?," "What was the perception of your workmate in the second phase of the game?," "What was the perception of your collaboration and degree of gratitude in the first phase of the game?," and "What was the perception of your collaboration and the degree of gratitude in the second phase of the game?." Participants could respond to items by assigning a Likert-scale score from 1 (perception of non-synchronicity/non-cooperation) to 3 (perception of synchrony and cooperation).

Autonomic Measures Recording and Analysis

The autonomic activity was recorded using two Expert2000 portable Biofeedback systems with a MULTI radio module (Schuhfried GmbH, Mödling, Austria) that allow to measure the level and response of SCL in μS and HR in beats per minute (bpm). The SCL value was recorded with an EDA1 gold electrode using current measurement at a sampling frequency of 2 kiloHertz (kHz). The use of alternating voltage prevents polarization. The measurement resolution for the SCL calculation is 12 nanosecond (ns) with a sampling frequency of 20 Hertz (Hz). HR was measured by the infrared absorption principle with a sampling frequency of 500 Hz. The parameter range was 30–200 bpm. Furthermore, the mobility of the non-dominant hand was monitored with an accelerometer in meter/square second (m/s2) integrated into the sending unit to ensure that the recordings were not compromised by hand movements. Trials with motor artifacts have been removed from the analyses.

DATA ANALYSIS

Questionnaire Responses

For the answers to the questionnaire, a preliminary analysis was conducted. Specifically, two mixed model ANOVAs with Block (pre vs. post) and Role (donor vs. receiver) as a repeated factor, and Condition as between-subjects factor (Cond: order 1 vs. order 2) were applied to questionnaire scores.

ANOVA results showed a significant effect for Block ($F[1,97] = 22.56$; $p < 0.0001$; $\eta^2 = 0.90$), with a greater perception of tuning after ($M = 3.0$; $SD = 0.07$) than before ($M = 1.04$; $SD = 0.05$) gift exchange. With regard to the perception of cooperation, instead, ANOVA showed a significant effect for Block ($F[1,97] = 34$; $p < 0.0001$; $\eta^2 = 0.94$), with a greater perception of cooperation later ($M = 2.99$; $DS = 0.06$) compared to before ($M = 1.34$; $SD = 0.03$) gift exchange.

Behavioral and Autonomic Data Analyses

Three main orders of analyses were performed: (1) Repeated measures ANOVAs on the modulation of the dependent variables (ACC, RTs, SCL, and HR) for the subjects throughout the task were conducted (A – intra-subjective analysis); (2) inter-subjects correlational indices were performed (B – inter-subjective analysis) to compute the synchronization values within each couple for each autonomic measure; and (3) repeated measures ANOVAs on such indices to assess differences in synchrony strength across the experimental conditions. For all the ANOVA tests, the degrees of freedom were corrected using Greenhouse-Geisser epsilon where appropriate, with the level of significance set at 0.05. Also, *post-hoc* comparisons (contrast analyses) were applied to the data. Bonferroni test was applied for multiple comparisons. In addition, the normality of the data distribution was preliminary tested (kurtosis and asymmetry tests). The normality assumption of the distribution was supported by these preliminary tests.

(A) Intra-Subjective Analysis

Behavioral Results

Accuracy and RTs scores were obtained for each subject using E-prime software during the three blocks. Specifically, for the analysis of ACC, the percentage of correct answers on the total answers was considered, while RTs were calculated starting from the presentation of the stimulus.

Before the pre-gift training condition, after the 120-s reference record, the subjects were provided a familiarization task. Specifically, two mixed-model ANOVAs were applied to ACC and RT with Blocks (1baseline vs. 2 vs. 3) as a repeated factor and Condition (Cond: order 1 vs. order 2) and Role (Role: donor vs. receiver) as between-subjects factor.

For ACC, ANOVA revealed a significant effect for Cond ($F[1,60] = 12.34$; $p < 0.001$; $\eta^2 = 0.38$), with better performance for order 1 than order 2; Block ($F[2,97] = 9.77$; $p < 0.001$; $\eta^2 = 0.32$) and Cond × Block ($F[2,97] = 12.98$; $p < 0.001$; $\eta^2 = 0.37$). In particular, as shown by *post-hoc* comparisons applied to interaction effects, order 1 revealed a higher ACC in block 2 ($F[1,31] = 9.32$; $p < 0.001$; $\eta^2 = 0.31$) and in block 3 more than block 1 ($F[1,31] = 8.87$; $p < 0.001$; $\eta^2 = 0.29$).

Moreover, block 3 differed from block 2 ($F[1,31] = 8.50$; $p < 0.001$; $\eta^2 = 0.29$) with higher ACC. In contrast, order 2 showed higher ACC only in block 3 more than block 1 ($F[1,31] = 9.08$; $p < 0.001$; $\eta^2 = 0.30$) and block 2 ($F[1,31] = 9.16$; $p < 0.001$; $\eta^2 = 0.32$; **Figure 2A**).

No significant effects were found for RTs.

Autonomic Measures

About SCL, Cond × Block interaction effect was significant ($F[2,97] = 7.94$, $p < 0.001$, $\eta^2 = 0.29$). Specifically, as revealed by *post-hoc* comparisons, there was an increase of SCL for

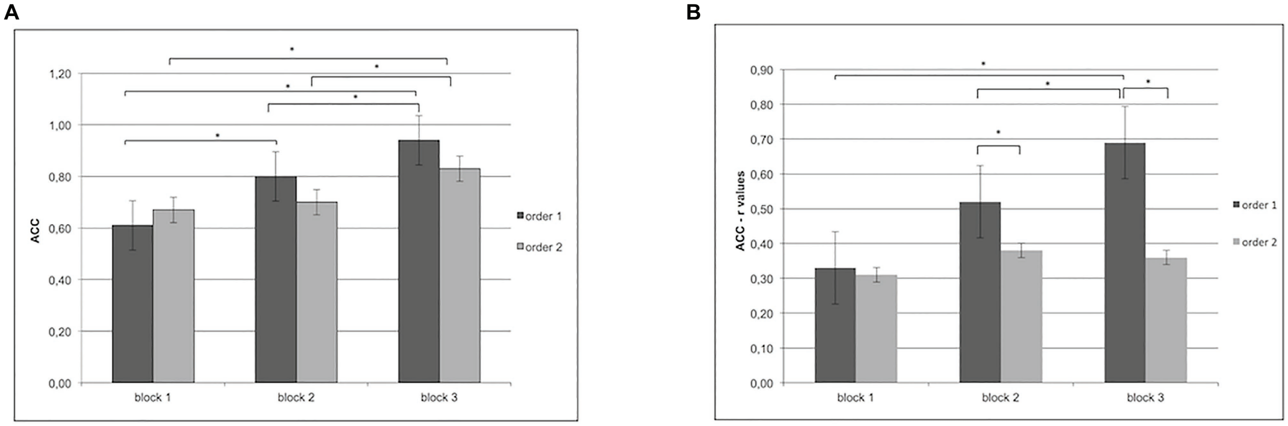


FIGURE 2 | **(A)** Increase in accuracy for order 1 in blocks 2 and 3 more than block 1 and in block 3 more than block 2. For order 2, instead, the figure shows an increase of accuracy in block 3 more than blocks 1 and 2. **(B)** Increase of inter-subjective accuracy (*r* values) for order 1 more than order 2 in block 2 and block 3. Moreover, for order 1, the figure shows an increase of inter-subjective accuracy (*r* values) in block 3 more than block 1. In addition, higher synchronicity was observed in order 1 for block 3 more than block 2.

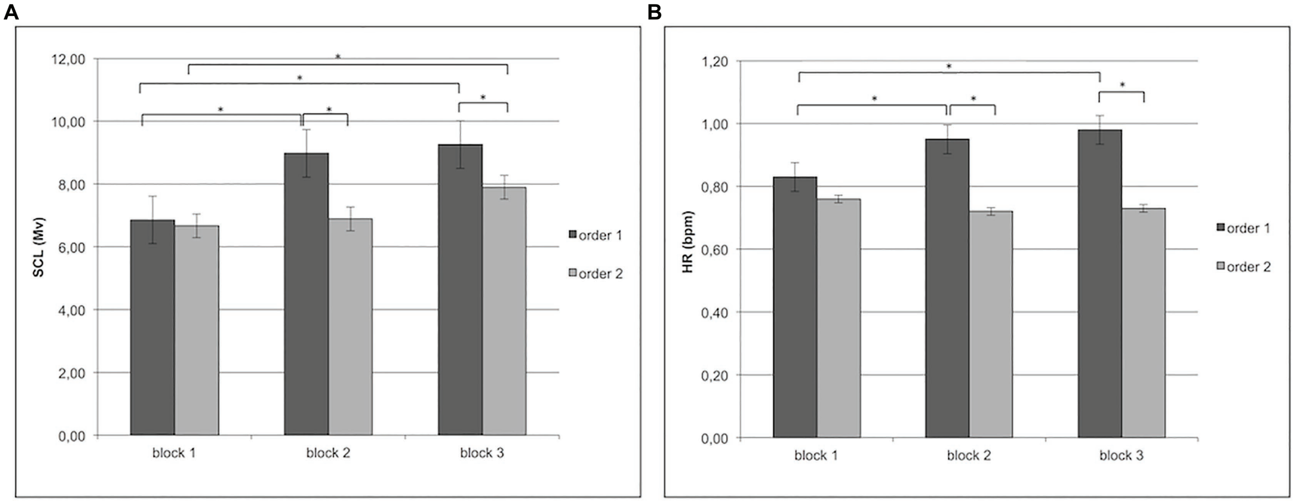


FIGURE 3 | **(A)** Increase in individuals' skin conductance level (SCL) for order 1 in block 2 and block 3 more than order 2. Moreover, the figure shows for order 1 an increase of SCL in blocks 2 and 3 more than block 1; while, for order 2, an increase of SCL in block 3 more than block 1. **(B)** For heart rate (HR), the figure shows an increase of HR for order 1 more than order 2 in block 2 and block 3. Specifically, for order 1, the figure shows an increase of HR in blocks 2 and 3 more than block 1.

order 1 more than order 2 in block 2 and block 3 (respectively, $F[1,31] = 9.56$, $p < 0.01$, $\eta^2 = 0.33$; $F[1,31] = 8.99$, $p < 0.01$, $\eta^2 = 0.29$). In addition, in order 1, block 2 and block 3 differed from block 1 with increased SCL ($F[1,31] = 8.55$, $p < 0.01$, $\eta^2 = 0.28$), whereas the differences in order 2 were found only between block 1 and block 3 (respectively, $F[1,31] = 8.90$, $p < 0.01$, $\eta^2 = 0.29$; $F[1,31] = 9.56$, $p < 0.01$, $\eta^2 = 0.32$; **Figure 3A**).

About HR, Cond × Block interaction effect was significant ($F[2,97] = 9.33$, $p < 0.001$, $\eta^2 = 0.35$). Specifically, as revealed by *post-hoc* comparisons, there was an increase of HR for order 1 more than order 2 in block 2 and block 3 (respectively, $F[1,31] = 12.09$, $p < 0.01$, $\eta^2 = 0.36$; $F[1,31] = 9.09$, $p < 0.01$, $\eta^2 = 0.31$). In addition, in order 1, block 2 and block 3 differed from block 1 with increased HR (respectively, $F[1,31] = 10.06$, $p < 0.01$, $\eta^2 = 0.37$; $F[1,31] = 9.50$, $p < 0.01$, $\eta^2 = 0.34$; **Figure 3B**).

(B) Inter-Subjective Analyses

Calculation of Correlational Indices

The synchronization indices were calculated by correlational coefficients (Pearson coefficients) applied to the data for each behavioral (ACC and RTs) and autonomic index (SCL and HR; Hernandez et al., 2014).

According to these indices, the subsequent step of analysis was finalized to test the statistical significance of independent factor Blocks (1baseline vs. 2 vs. 3) and Condition (Cond: order 1 vs. order 2) on these correlational indices for each couple by using repeated measures ANOVAs.

Behavioral Measures

For ACC, ANOVA showed a significant Cond × Block interaction effect ($F[2,97] = 8.89$, $p < 0.01$, $\eta^2 = 0.29$). Specifically, pairwise *post-hoc* comparisons revealed a significant higher synchronization (higher r values) for order 1 more than order 2 in block 2 and block 3 (respectively, $F[1,13] = 9.11$, $p < 0.01$, $\eta^2 = 0.32$; $F[1,13] = 8.44$, $p < 0.01$, $\eta^2 = 0.30$). In addition higher synchronicity was observed in order 1 for block 3 more than block 1 ($F[1,13] = 6.34$, $p < 0.01$, $\eta^2 = 0.26$) and block 2 ($F[1,13] = 6.89$, $p < 0.01$, $\eta^2 = 0.27$; **Figure 2B**).

No significant effect was found for RTs.

Autonomic Measures

For what concerns SCL coefficient data, no significant differences in synchrony were found for the main or interaction effects.

In contrast, significant main effect for Condition × Block was found for HR ($F[1,15] = 8.87$, $p < 0.01$, $\eta^2 = 0.29$), which showed increased HR synchrony for order 1 more than order 2 in block 2 ($F[1,13] = 6.34$, $p < 0.01$, $\eta^2 = 0.26$) and block 3 ($F[1,13] = 6.34$, $p < 0.01$, $\eta^2 = 0.26$). In addition in order 1, increased synchrony was found more in block 2 ($F[1,13] = 6.34$, $p < 0.01$, $\eta^2 = 0.26$) and block 3 ($F[1,13] = 7.09$, $p < 0.01$, $\eta^2 = 0.27$) than block 1, whereas in order 2, increased synchronicity was found only for block 3 compared to block 1 ($F[1,13] = 6.34$, $p < 0.01$, $\eta^2 = 0.26$; **Figure 4**).

DISCUSSION

The present research has analyzed a joint action dynamic through the application of a multimethodological paradigm requiring the recording of participants' autonomic responses and behavioral performance during a cooperative activity involved a beginning or middle gift exchange. In addition to previous studies (Balconi et al., 2019a,b, 2020; Balconi and Fronda, 2020) aimed at detecting central indices (EEG or hemodynamic activity), which allow obtaining different information on different cognitive processes (Herrmann et al., 2003, 2008), the measurement of autonomic activity can be considered as a potential biological marker of emotions (Tupak et al., 2014), which allows to better define the interaction between peripheral and central systems (Furmark et al., 1997; Lang et al., 2000). Data were analyzed both at the individual (intra-subject) and at the dyadic level (inter-subject) to calculate PS and before and after gift exchange. Indeed, we hypothesized that sharing a gift could be accompanied by a positive emotional engagement that could influence both the behavioral and the physiological responses.

The analyses allowed to highlight some main results. (1) An improvement of both behavioral and autonomic responsivity has emerged in order 1, when gift donation came earlier, compared to a delayed exchange. A similar pattern emerged also at the inter-personal level, with increased physiological synchronization after the early gift. (2) The modulation of such effects by block variable was observed. In detail, both at the individual and the joint level, the advantage of order 1 was maximized over the blocks in

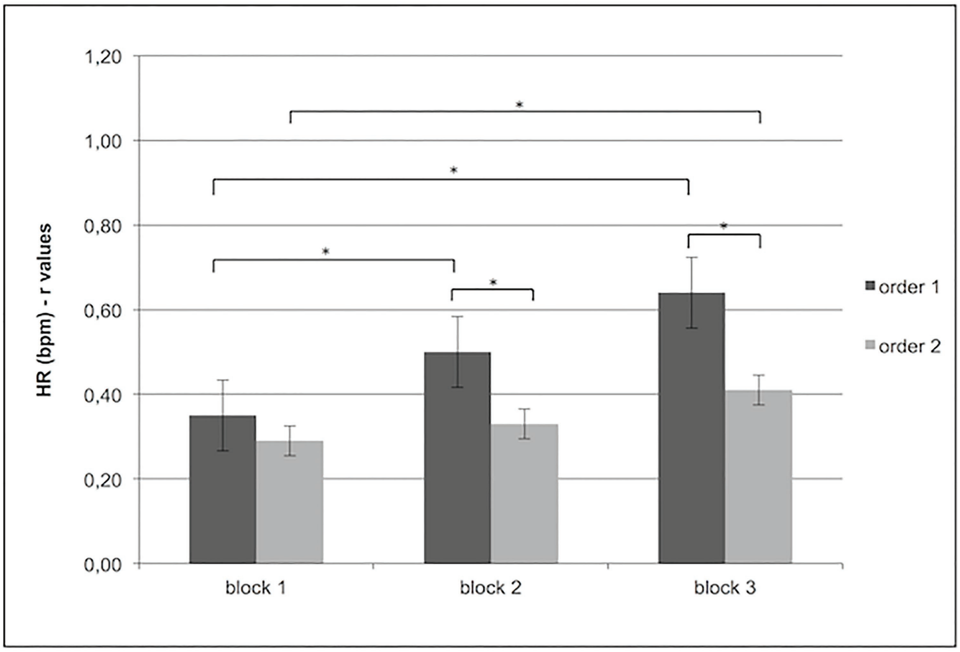


FIGURE 4 | Increased HR subjects' synchrony (r values) for order 1 more than order 2 in block 2 and block 3. Moreover, for order 1, the figure shows an increase in HR synchrony in blocks 2 and 3 more than block 1 and in order 2, an increase of HR subjects' synchrony was found block 3 more than block 1.

favor of the last block. (3) Both cardiovascular (HR) and electrodermal (SCL) indices proved to be effective in detecting participants' individual and joint responses. However, HR was more sensitive to such dynamics, especially at the interpersonal level. Finally, (4) although condition and blocks were able to modulate participants' responses during the task, the specific role adopted (donor or receiver) was not influential.

Beginning with the first point (1), an earlier gift exchange was associated with increased accuracy and higher autonomic responses in detecting the attentional targets both at the individual level and the interpersonal level. Such result can be interpreted by referring to the construct of gratitude. In fact, giving and receiving a gift within an interpersonal interaction is associated with a positive feedback loop which leads to emotional sharing (Aknin et al., 2011). Previous research, indeed, already underlined the influence of gratitude and positive emotions in strengthening cooperative behavior (Rumble et al., 2010). Such greater tuning was also explicitly attested by the participants through the self-report questionnaire. According to previous research, these emotional states can be the consequence of sharing a pleasant experience, which increases the feeling of being part of a whole and the sense of interpersonal cohesion (Balconi and Pagani, 2015; Chung et al., 2015; Vanutelli et al., 2017). Moreover, such result is of particular importance since this kind of interpersonal attuning makes the implementation of prosocial behaviors more likely (Ruby and Decety, 2004; Spinella, 2005; Balconi et al., 2011; Balconi and Bortolotti, 2012) and represents a social glue thanks to a reciprocity mechanism (Balconi and Lucchiari, 2006; Balconi and Pozzoli, 2007).

However, to better understand our results, it is also important to consider trust mechanisms. Indeed, although both order 1 and order 2 involve gift exchange, order 1 could have facilitated bond construction, emotional sharing, and cooperation, since it immediately allowed participants to increase their trust based on initial gift exchange. Previous research underlined how the experience of trust is highly influenced by mood and emotions (Jones and George, 1998). Accordingly, we believe that an interaction based on trust since its beginning could be more efficient in engaging participants in a cooperative activity. Such result finds support in previous research that showed a relation between bond construction, cooperation, and interpersonal coordination, even in studies with unrelated participants (Chung et al., 2015; Balconi and Vanutelli, 2016; Vanutelli et al., 2017; Balconi et al., 2018). For example, research on autonomic synchrony showed that the covariation between couples' physiological indices can reveal insights about the quality of their interaction representing a key marker of social engagement (Vanutelli et al., 2017).

This hypothesis is also corroborated by the second result (2). Indeed, the advantage of an earlier gift, shown by significant effects founded across the three blocks is potentially relevant to suppose a trend with an increased synchronization during the task and over the time. Even if order 2 displays an increasing synchronization over blocks, order 1 has a more definite distribution of both behavioral and autonomic responses, which can be appreciated especially between blocks 2 and 3. Indeed, in order 1, the gift is donated after the first block. Thus, there is sufficient time for the participants to develop dyadic strategies that become refined and improved in the last block. Contrarily, in order 2, the gift is donated between blocks 2 and 3, in a way that the effect of positive emotions cannot be translated into consolidated joint strategies both at behavioral and autonomic levels, as shown for order 1.

For what concerns this last point, it is important to underline the functional role of the autonomic responses (3). At the individual level, both SCL and HR proved to be sensitive to gift exchange modulation. However, referring to PS, a systematic synchronization emerged only for HR correlational indices. Interestingly, although SCL can be interpreted in relation to emotional arousal (Sohn et al., 2004), an increase in HR indices has been reported when experiencing highly positive emotional states (Ekman et al., 1983; Levenson, 1992; Sinha et al., 1992). Also, the covariation of such indices within an interpersonal relationship can be an index of emotional closeness (Konvalinka et al., 2010) and trust. Moreover, the degree of HR synchrony could predict participants' expectations about the moves of their partners (Mitkidis et al., 2015). Since SCL is generally interpreted as an emotional-arousal related measure (Picard et al., 2016), the simple increasing of arousal is not enough to support a synchronous response by the members of the dyad.

Finally, the role played by the two members of the dyad (donor or receiver) did not have any significant effects in modulating participants' responses. Thus, we could assume that both the act of giving and receiving a gift can similarly contribute to creating stronger cooperative ties, underlining how the implementation of prosocial behaviors can represent a social reward, even without a material return (Vanutelli et al., 2016). However, the absence of significant differences based on the role may be due to the familiarity and previous friendship between donor and receiver.

CONCLUSION

To conclude, the present study showed how gratitude and trust elicited within a gift donation protocol can modify the individuals' behavioral and autonomic responses. The moment of gift exchange can actually influence the creation of a positive emotional feedback loop and the perception of trust. When the gift is donated early, participants have the possibility to build a safe interpersonal space to develop a shared cognition, which provide an increase of performance, autonomic responses, and synchrony. This hypothesis is supported by the fact that this pattern proved to follow an exponential path, which reaches its maximum strength in the last experimental block, and by the specific autonomic index involved.

We believe that our paradigm is innovative and that it could be shared among scholars and applied to different real-life contexts where trust can be improved to provide a better emotional experience, such as the organizational, the educational, and the clinical framework. Some further developments may be proposed: The integration of neurophysiological and neuroimaging measurements would allow obtaining information on the neural individuals' synchronization and the integration of psychometric measures would allow evaluating some personalities traits; the comparison between dyad of subjects with and without previous friendship to better explore the effect of familiarity in distinguishing the role effect. Finally, the role of gender in cooperation and trust could be explored.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by local Ethics Committee of the Department of Psychology of the Catholic University of Milan. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MB contributed to the conception and design of the study and wrote the first draft and each section of the manuscript. MB and GF contributed to manuscript final writing and revision, and read and approved the submitted version.

REFERENCES

Adolphs, R. (2003). Cognitive neuroscience: cognitive neuroscience of human social behavior. *Nat. Rev. Neurosci.* 4, 165–178. doi: 10.1038/nrn1056

Aknin, L. B., Sandstrom, G. M., Dunn, E. W., and Norton, M. I. (2011). It's the recipient that counts: spending money on strong social ties leads to greater happiness than spending on weak social ties. *PLoS One* 6:e17018. doi: 10.1371/journal.pone.0017018

Algoe, S. B., Haidt, J., and Gable, S. L. (2008). Beyond reciprocity: gratitude and relationships in everyday life. *Emotion* 8, 425–429. doi: 10.1037/1528-3542.8.3.425

Balconi, M., and Bortolotti, A. (2012). Detection of the facial expression of emotion and self-report measures in empathic situations are influenced by sensorimotor circuit inhibition by low-frequency rTMS. *Brain Stimul.* 5, 330–336. doi: 10.1016/j.brs.2011.05.004

Balconi, M., and Bortolotti, A. (2013). Emotional face recognition, empathic trait (BEES), and cortical contribution in response to positive and negative cues. The effect of rTMS on dorsal medial prefrontal cortex. *Cogn. Neurodyn.* 7, 13–21. doi: 10.1007/s11571-012-9210-4

Balconi, M., Bortolotti, A., and Gonzaga, L. (2011). Emotional face recognition, EMG response, and medial prefrontal activity in empathic behavior. *Neurosci. Res.* 71, 251–259. doi: 10.1016/j.neures.2011.07.1833

Balconi, M., and Canavesio, Y. (2013). Emotional contagion and trait empathy in prosocial behavior in young people: the contribution of autonomic (facial feedback) and Balanced Emotional Empathy Scale (BEES) measures. *J. Clin. Exp. Neuropsychol.* 35, 41–48. doi: 10.1080/13803395.2012.742492

Balconi, M., and Canavesio, Y. (2014). High-frequency rTMS on DLPFC increases prosocial attitude in case of decision to support people. *Soc. Neurosci.* 9, 82–93. doi: 10.1080/17470919.2013.861361

Balconi, M., and Fronda, G. (2020). The "gift effect" on functional brain connectivity. Inter-brain synchronization when prosocial behavior is in action. *Sci. Rep.* 10:5394. doi: 10.1038/s41598-020-62421-0

Balconi, M., Fronda, G., and Vanutelli, M. E. (2019a). A gift for gratitude and cooperative behavior: brain and cognitive effects. *Soc. Cogn. Affect. Neurosci.* 14, 1317–1327. doi: 10.1093/scan/nsaa003

Balconi, M., Fronda, G., and Vanutelli, M. E. (2019b). Donate or receive? Social hyperscanning application with fNIRS. *Curr. Psychol.* 38, 991–1002. doi: 10.1007/s12144-019-00247-4

Balconi, M., Fronda, G., and Vanutelli, M. E. (2020). When gratitude and cooperation between friends affect inter-brain connectivity for EEG. *BMC Neurosci.* 21:14. doi: 10.1186/s12868-020-00563-7

Balconi, M., Gatti, L., and Vanutelli, M. E. (2018). Cooperate or not cooperate EEG, autonomic, and behavioral correlates of ineffective joint strategies. *Brain Behav.* 8:e00902. doi: 10.1002/brb3.902

Balconi, M., and Lucchiari, C. (2006). EEG correlates (event-related desynchronization) of emotional face elaboration: a temporal analysis. *Neurosci. Lett.* 392, 118–123. doi: 10.1016/j.neulet.2005.09.004

Balconi, M., and Pagani, S. (2015). Social hierarchies and emotions: cortical prefrontal activity, facial feedback (EMG), and cognitive performance in a dynamic interaction. *Soc. Neurosci.* 9, 82–93. doi: 10.1080/17470919.2014.977403

Balconi, M., and Pozzoli, U. (2007). Event-related oscillations (EROs) and event-related potentials (ERPs) comparison in facial expression recognition. *J. Neuropsychol.* 1, 283–294. doi: 10.1348/174866407X184789

Balconi, M., and Vanutelli, M. E. (2016). Competition in the brain. The contribution of EEG and fNIRS modulation and personality effects in social ranking. *Front. Psychol.* 7:1587. doi: 10.3389/fpsyg.2016.01587

Balconi, M., and Vanutelli, M. E. (2017a). Cooperation and competition with hyperscanning methods: review and future application to emotion domain. *Front. Comput. Neurosci.* 11:86. doi: 10.3389/fncom.2017.00086

Balconi, M., and Vanutelli, M. E. (2017b). Interbrains cooperation: hyperscanning and self-perception in joint actions. *J. Clin. Exp. Neuropsychol.* 39, 607–620. doi: 10.1080/13803395.2016.1253666

Barrett, L. F. (2006). Are emotions natural kinds? *Perspect. Psychol. Sci.* 1, 28–58. doi: 10.1111/j.1745-6916.2006.00003.x

Boucsein, W. (2012). *Electrodermal activity. 2nd Edn*. New York: Springer Science + Business Media.

Butler, J. (2011). *Bodies that matter: On the discursive limits of sex*. London: Routledge.

Chaspari, T., Baucom, B., Timmons, A. C., Tsiartas, A., Del Piero, L. B., Baucom, K. J. W., et al. (2015). "Quantifying EDA synchrony through joint sparse representation: a case-study of couples' interactions" in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP);* April 19–24, 2015; 62, 960–971.

Chauhan, B., Mathias, C. J., and Critchley, H. D. (2008). Autonomic contributions to empathy: evidence from patients with primary autonomic failure. *Auton. Neurosci. Basic Clin.* 140, 96–100. doi: 10.1016/j.autneu.2008.03.005

Chung, D., Yun, K., and Jeong, J. (2015). Decoding covert motivations of free riding and cooperation from multi-feature pattern analysis of EEG signals. *Soc. Cogn. Affect. Neurosci.* 10, 1210–1218. doi: 10.1093/scan/nsv006

Diamond, L. M., and Otter-Henderson, K. D. (2007). "Physiological measures" in *Handbook of research methods in personality psychology*. eds. R. W. Robins, R. C. Fraley and R. F. Krueger (New York: The Guilford Press), 370–388.

Duclos, R., and Barasch, A. (2014). Prosocial behavior in intergroup relations: how donor self-construal and recipient group-membership shape generosity. *J. Consum. Res.* 41, 93–108. doi: 10.1086/674976

Dumas, G., Lachat, F., Martinerie, J., Nadel, J., and George, N. (2011). From social behavior to brain synchronization: Review and perspectives in hyperscanning. *IRBM* 32, 48–53. doi: 10.1016/j.irbm.2011.01.002

Ekman, P., Levenson, R. W., and Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science* 221, 1208–1210. doi: 10.1126/science.6612338

Flynn, F. J., and Brockner, J. (2003). It's different to give than to receive: predictors of givers' and receivers' reactions to favor exchange. *J. Appl. Psychol.* 88, 1034–1045. doi: 10.1037/0021-9010.88.6.1034

Froh, J. J., Sefick, W. J., and Emmons, R. A. (2008). Counting blessings in early adolescents: an experimental study of gratitude and subjective well-being. *J. Sch. Psychol.* 46, 213–233. doi: 10.1016/j.jsp.2007.03.005

Furmark, T., Fischer, H., Wik, G., Larsson, M., and Fredrikson, M. (1997). The amygdala and individual differences in human fear conditioning. *Neuroreport* 8, 3957–3960. doi: 10.1097/00001756-199712220-00021

Giuliano, R. J., Skowron, E. A., and Berkman, E. T. (2015). Growth models of dyadic synchrony and mother-child vagal tone in the context of parenting at-risk. *Biol. Psychol.* 105, 29–36. doi: 10.1016/j.biopsycho.2014.12.009

Golland, Y., Arzouan, Y., and Levit-Binnun, N. (2015). The mere co-presence: synchronization of autonomic signals and emotional responses across co-present individuals not engaged in direct interaction. *PLoS One* 10:e0125804. doi: 10.1371/journal.pone.0125804

Gray, J. R. (2004). Integration of emotion and cognitive control. *Curr. Dir. Psychol. Sci.* 13, 46–48. doi: 10.1111/j.0963-7214.2004.00272.x

Guastello, S. J., Pincus, D., and Gunderson, P. R. (2006). Electrodermal arousal between participants in a conversation: nonlinear dynamics and linkage effects. *Nonlinear Dynamics Psychol. Life Sci.* 10, 365–399.

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., and Keysers, C. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends Cogn. Sci.* 16, 114–121. doi: 10.1016/j.tics.2011.12.007

Hatfield, E., Cacioppo, J. T., and Rapson, R. L. (1994). *Emotional contagion: Cambridge studies in emotion and social interaction*. Cambridge: Cambridge University Press.

Hernandez, J., Riobo, I., Rozga, A., Abowd, G. D., and Picard, R. W. (2014). "Using electrodermal activity to recognize ease of engagement in children during social interactions" in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*; September 13–17, 2014; 307–317.

Herrmann, M. J., Ehlis, A. C., and Fallgatter, A. J. (2003). Prefrontal activation through task requirements of emotional induction measured with NIRS. *Biol. Psychol.* 64, 255–263. doi: 10.1016/S0301-0511(03)00095-4

Herrmann, M. J., Huter, T., Plichta, M. M., Ehlis, A. C., Alpers, G. W., Mühlberger, A., et al. (2008). Enhancement of activity of the primary visual cortex during processing of emotional stimuli as measured with event-related functional near-infrared spectroscopy and event-related potentials. *Hum. Brain Mapp.* 29, 28–35. doi: 10.1002/hbm.20368

Isen, A. M. (2009). "A role for neuropsychology in understanding the facilitating influence of positive affect on social behavior and cognitive processes" in *The Oxford handbook of positive psychology. 2nd Edn*. eds. C. R. Snyder and S. Lopez (New York: Oxford University Press), 503–518.

Jones, G. R., and George, J. M. (1998). The experience and evolution of trust: implications for cooperation and teamwork. *Acad. Manag. Rev.* 23, 531–546. doi: 10.5465/amr.1998.926625

Koike, T., Tanabe, H. C., and Sadato, N. (2015). Hyperscanning neuroimaging technique to reveal the "two-in-one" system in social interactions. *Neurosci. Res.* 90, 25–32. doi: 10.1016/j.neures.2014.11.006

Konvalinka, I., Vuust, P., Roepstorff, A., and Frith, C. D. (2010). Follow you, follow me: continuous mutual prediction and adaptation in joint tapping. *Q. J. Exp. Psychol.* 108, 8514–8519. doi: 10.1080/17470218.2010.497843

Konvalinka, I., Xygalatas, D., Bulbulia, J., Schjodt, U., Jegindo, E. -M., Wallot, S., et al. (2011). Synchronized arousal between performers and related spectators in a fire-walking ritual. *Proc. Natl. Acad. Sci. U. S. A.* 63, 2220–2230. doi: 10.1073/pnas.1016955108

Lambert, N. M., Clark, M. S., Durtschi, J., Fincham, F. D., and Graham, S. M. (2010). Benefits of expressing gratitude: expressing gratitude to a partner changes one's view of the relationship. *Psychol. Sci.* 21, 574–580. doi: 10.1177/0956797610364003

Lang, P. J., Davis, M., and Öhman, A. (2000). Fear and anxiety: animal models and human cognitive psychophysiology. *J. Affect. Disord.* 61, 137–159. doi: 10.1016/S0165-0327(00)00343-8

Levenson, R. W. (1992). Autonomic nervous system differences among emotions. *Psychol. Sci.* 3, 23–27. doi: 10.1111/j.1467-9280.1992.tb00251.x

Levenson, R. W., and Ruef, A. M. (1992). Empathy: a physiological substrate. *J. Pers. Soc. Psychol.* 63, 234–246. doi: 10.1037/0022-3514.63.2.234

Marci, C. D., Ham, J., Moran, E., and Orr, S. P. (2007). Physiologic correlates of perceived therapist empathy and social-emotional process during psychotherapy. *J. Nerv. Ment. Dis.* 195, 103–111. doi: 10.1097/01.nmd.0000253731.71025.fc

McCullough, M. E., and Tsang, J. A. (2004). "Parent of the virtues? The prosocial contours of gratitude" in *Series in affective science. The psychology of gratitude*. eds. R. A. Emmons and M. E. McCullough (New York: Oxford University Press), 123–141.

Mick, D. G., and Demoss, M. (1990). Self-gifts: phenomenological Insights from four contexts. *J. Consum. Res.* 17, 322–332. doi: 10.1086/208560

Mick, D. G., and Faure, C. (1998). Consumer self-gifts in achievement contexts: the role of outcomes, attributions, emotions, and deservingness. *Int. J. Res. Mark.* 15, 293–307. doi: 10.1016/S0167-8116(98)00006-8

Mirgorodsky, V. I., Dementienko, V. V., Gerasimov, V. V., Peshin, S. V., Dorokhov, V. B., and Shakhnarovich, V. M. (2013). Correlations between individual features of electrodermal response during viewing of the same videos. *Int. J. Neurosci. Behav. Sci.* 1, 5–12. doi: 10.13189/ijnbs.2013.010102

Mitkidis, P., McGraw, J. J., Roepstorff, A., and Wallot, S. (2015). Building trust: heart rate synchrony and arousal during joint action increased by public goods game. *Physiol. Behav.* 1, 101–106. doi: 10.1016/j.physbeh.2015.05.033

Müller, V., and Lindenberger, U. (2011). Cardiac and respiratory patterns synchronize between persons during choir singing. *PLoS One* 6:e24893. doi: 10.1371/journal.pone.0024893

Nadler, R. T., Rabi, R., and Minda, J. P. (2010). Better mood and better performance: learning rule-described categories is enhanced by positive mood. *Psychol. Sci.* 21, 1770–1776. doi: 10.1177/0956797610387441

Niedenthal, P. M. (2007). Embodying emotion. *Science* 316, 1002–1005. doi: 10.1126/science.1136930

Penner, L. A., Dovidio, J. F., Piliavin, J. A., and Schroeder, D. A. (2005). Prosocial behavior: multilevel perspectives. *Annu. Rev. Psychol.* 56, 365–392. doi: 10.1146/annurev.psych.56.091103.070141

Picard, R. W., Fedor, S., and Ayzenberg, Y. (2016). Multiple arousal theory and daily-life electrodermal activity asymmetry. *Emot. Rev.* 8, 62–75. doi: 10.1177/1754073914565517

Richardson, D. C., Dale, R., and Kirkham, N. Z. (2007). The art of conversation is coordination: common ground and the coupling of eye movements during dialogue: Research article. *Psychol. Sci.* 18, 407–413. doi: 10.1111/j.1467-9280.2007.01914.x

Ruby, P., and Decety, J. (2004). How would you feel versus how do you think she would feel? A neuroimaging study of perspective-taking with social emotions. *J. Cogn. Neurosci.* 16, 988–999. doi: 10.1162/0898929041502661

Rumble, A. C., van Lange, P. A. M., and Parks, C. D. (2010). The benefits of empathy: when empathy may sustain cooperation in social dilemmas. *Eur. J. Soc. Psychol.* 40, 856–866. doi: 10.1002/ejsp.659

Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., et al. (2010). Minds made for sharing: Initiating joint attention recruits reward-related neurocircuitry. *J. Cogn. Neurosci.* 22, 2702–2715. doi: 10.1162/jocn.2009.21401

Sinha, R., Lovallo, W. R., and Parsons, O. A. (1992). Cardiovascular differentiation of emotions. *Psychosom. Med.* 54, 422–435. doi: 10.1097/00006842-199207000-00005

Sohn, J. H., Sokhadze, E., and Watanuki, S. (2004). Electrodermal and cardiovascular manifestations of emotions in children. *J. Physiol. Anthropol. Appl. Hum. Sci.* 20, 55–64. doi: 10.2114/jpa.20.55

Spinella, M. S. (2005). Prefrontal substrates of empathy: psychometric evidence in a community sample. *Biol. Psychol.* 70, 175–181. doi: 10.1016/j.biopsycho.2004.01.005

Tupak, S. V., Dresler, T., Guhn, A., Ehlis, A. C., Fallgatter, A. J., Pauli, P., et al. (2014). Implicit emotion regulation in the presence of threat: neural and autonomic correlates. *NeuroImage* 85, 372–379. doi: 10.1016/j.neuroimage.2013.09.066

Vanutelli, M. E., Gatti, L., Angioletti, L., and Balconi, M. (2017). Affective synchrony and autonomic coupling during cooperation: a hyperscanning study. *Biomed. Res. Int.* 2017:3104564. doi: 10.1155/2017/3104564

Vanutelli, M. E., Gatti, L., Angioletti, L., and Balconi, M. (2018). May the best joint-actions win: physiological linkage during competition. *Appl. Psychophysiol. Biofeedback* 43, 227–237. doi: 10.1007/s10484-018-9402-8

Vanutelli, M. E., Nandrino, J. -L., and Balconi, M. (2016). The boundaries of cooperation: sharing and coupling from ethology to neuroscience. *Neuropsychol. Trends* 19, 83–104. doi: 10.7358/neur-2016-019-vanu

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

OPINION
published: 24 August 2021
doi: 10.3389/fpsyg.2021.560653

First Selectively Imitate Responses, Then Associate With Function

Elpida Tzafestas [1,2]*

[1] *Laboratory of Cognitive Science, Department of History and Philosophy of Science, National and Kapodistrian University of Athens, Athens, Greece,* [2] *Department of Electrical and Computer Engineering, Technische Universität München, München, Germany*

Keywords: imitation, association, selection, communication, social learning, behavioral prediction

Edited by:
Andreas K. Engel,
University Medical Center
Hamburg-Eppendorf, Germany

Reviewed by:
Desmond Mulligan,
University of British Columbia, Canada

***Correspondence:**
Elpida Tzafestas
etzafestas@phs.uoa.gr

Specialty section:
This article was submitted to
Cognitive Science,
a section of the journal
Frontiers in Psychology

Received: *09 May 2020*
Accepted: *28 June 2021*
Published: *24 August 2021*

Citation:
Tzafestas E (2021) First Selectively
Imitate Responses, Then Associate
With Function.
Front. Psychol. 12:560653.
doi: 10.3389/fpsyg.2021.560653

INTRODUCTION

Imitation is ubiquitous: human adults imitate, human infants, preschoolers, and adolescents all imitate, some animals too imitate. Moreover, imitation may be automatic, spontaneous, or deliberate and humans may specifically select their imitation models as well as the imitated target feature or object. Finally, language, gestures, motor patterns, and high-level behaviors can all be imitated at varying degrees of detail and in a variety of modulating contexts. These are some major aspects of the study of the broad class of cognitive processes called imitation, whether and how it has evolved and its relation to empathy, mind-reading, language, culture, and social learning in general (Hurley and Chater, 2005; Heyes et al., 2009; Jones, 2009, Shea, 2009; Claidière and Sperber, 2010; Gerrans, 2013; Hodges, 2014). Because of this ubiquity, imitation is sometimes used with diversified or more restricted meanings, such as mimicry and emulation (van Baaren et al., 2009; Whiten et al., 2009), and research efforts often center on the neural and behavioral organization that allows imitation to happen (Brass and Heyes, 2005).

We are focusing on the relation of imitation with association, as a basic mechanism of behavioral emergence, and with communication, as the general function of imitation (communication need not be thought as purely linguistic, though[1]). Our goal is to discuss the nexus of the three concepts (imitation, association, communicative function) and to propose a plausible view of the initial steps taken by evolution in the development of general social and communicative behavior. Thus, we propose an incremental process of generative and selective imitation starting from proto-imitation that replicates external signals without associating with target objects or functionality, and proceeding to proto-association that relates to higher-order imitation and attributes "meaning" or function to external signals. We are therefore discussing a model that allows proto-imitation outside explicit communication but may allow emergence of communication in the medium or long term ontogenetically. The idea is to separate response imitation from response association to external meaning so as to make room both for species that can imitate but not associate meaning and for disabled humans that cannot associate well or even proto-imitate well. Thus, we regard and model association as a generic functional concept, initially Hebbian at the neural level ("what fires together, wires together"), but more intricate at higher levels (Cooper et al., 2013).

In what follows, we present in order the proto-imitation concept and model, the predictions about its function in regular configurations or in cognitively extreme cases, the basic meaning association model and, lastly, further behavioral high-level predictions at organism level. We have verified computationally some of the predictions elsewhere and all of them match actual evidence obtained experimentally. We conclude with further thoughts about the proposed model.

---

[1] Imitation at any level has always a communicative value, in the sense that what is imitated really is or is expected to be functional in the environment, where function can be anything from immediate survival to social bond construction to human language learning. Like any other behavioral shaping mechanism, however, imitation may in the end fail to perform as expected.

SELECTIVE IMITATION

We are adopting the view that the ontegenetic development at the neural level follows the same principles as Darwinian evolution at the population level (Edelman, 1987) and that any novel responses should be generated internally and selected within the environment rather than be directly "instructed" by it. Thus, an organism can express a number of intrinsically encoded responses that can be produced spontaneously or automatically or unintentionally (cf. Heyes, 2011 and "reflex practicing and conditioning," Piaget, 1947/2001) rather than computed rationally from the top to down (Gergely et al., 2002) and these can correspond loosely to real sensory or neural patterns such as vocal parameters. We call these responses "eigen-frequencies" or "eigen-responses," on the one hand as an engineering term that shows a value that can excite the organism to produce a response, and on the other hand with an eye to a coupled neural oscillators or an entrainment implementation (Buzsáki, 2006; Ansermin et al., 2016). These responses can be modeled as dependent on a recognition or excitation threshold (T) and have accordingly varying degrees of affinity to a given signal. The response to an external signal is the eigenresponse with the highest affinity. At each step, new diversified eigenresponses emerge proportionally to the affinity of the previous ones. The highest matching responses reproduce massively, while the lowest ones vanish and are replaced by newly generated random eigenresponses. An exploration factor (E) is also necessary, which is the rate of random eigenresponse replacement independently of affinity. The overall affinity to an external signal is the average affinity of all the responses, thus it is internally generated and not externally imposed/designed/taught in any way. The internal evaluation through affinity makes the model selective rather than instructive, but we would gladly count shaping within a constant environment as indirect teaching. The affinity measure expresses *how "well" an organism recognizes and can reproduce (imitate) a signal* and may therefore serve as a basis for subsequent emergence of communication, when meaning is introduced to the interaction with the environment. The actual speed of imitation/learning and self-organization outcome in a given environment depend on both the organism's eigenresponse repertoire and setup and on the dynamics of stimulation by the environment. **Figure 1** presents a general functional architecture that can support our model and how it makes sense from an evolution standpoint.

BEHAVIORAL EXTREMES

Our model uses a number of cognitive parameters that have a "normal" range that we expect to have been tuned by biological evolution. We can predict that extreme setups will still be residually present in a population of organisms and will lead to extreme results that correspond to either behavioral deficits or exceptional performance. More specifically, if the set of eigenresponses is too small, the organism may not be able to imitate and develop the whole set of responses necessary for efficient long-term survival. Such an organism may appear as cognitively impaired and slow in learning or incapable of it. Similar results are predicted when the recognition thresholds (T) are too low, in which case spontaneous response matching will be rare to start with. A final complication is when the response exploration factor (E) is too low and the proper response cannot

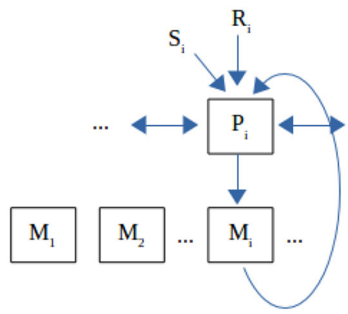


FIGURE 1 | Functional organization of the imitation system. **(1) Affinity.** A signal S_i is initially perceived as a sensory pattern by a perceptual neural component P_i that elicits an originally random motor response M_j, whose output can also be perceived by the same system. If responses that have a communicative value are selected by the environment (like for example a phonetic response from M_i matching the phonetic input S_i, such as an "a" pronounced by self being perceived as an "a" pronounced by another) or if there appear P_i's that happen to match S_i and M_i output, then evolution will favor the emergence of the pathways of the type P_i-M_i, where the output of M_i is perceived by P_i as the closest to S_i among all M_j. The pathway will show as higher population mass with larger total activation volume and diversity within the population, thus corresponding to our model of a "highly reproducing eigen-response." **(2) Self-organization**. None of P_i, M_i are born as invariants. All of them are subject to continuous reorganization, where P_i strives to better discriminate and assimilate (the closest) S_i, M_i strives to co-stabilize with P_i, etc. Moreover, there will be lateral interactions between P_i's and between M_i's because in principle P_i's overlap neurally to a degree, and so do M_i's. Thus, it takes some self-organization effort for each of them to stabilize to a discrete functional role in the system interfering the least with the other components. Again, with all these in place, evolution will favor the emergence of initial ("innate") structures that can self-organize correctly and with little effort given consistent stimulation. Exploration is the process by which P_i's and M_i's change intrinsically. **(3) Higher-level association.** External signals are not just reproduced by the motor system, but they also refer to external objects (S_i refers to R_i). Association of S_i's with R_i's will again pass through the organism's perceptual system (*via* a similar system as the previous level, but abstractly shown in the figure), so that either of them can finally trigger the same response M_i. We note that this is a minimal mirroring property, but insufficient to produce "intention understanding" or anything cognitively more advanced. Structurally (meaning) association is a process parallel to the previous ones and the correct associations could emerge independently of whether the proper P_i-M_i pathways have been built. But they are much easier to emerge when these pathways are in place, because then the S_i-P_i-M_i route may guide narrowing the scope of reference to the correct R_i. **(4) Representations.** There is no such thing in this organization, because neither perceptual nor motor systems represent anything real. However, when an organism learns to act consistently on external input, it acts *as if* representations were present (but, we are aware that for some thinkers systematic consistency and an AS-IF representation IS a true representation; we accept this as valid common ground for communication between "computationalists" and "connectionists," but see discussion by Chemero, 2009). **(5) Comments.** This model is similar to the proposal of Pickering and Garrod (2014) for language, except that unlike it there is selection instead of explicit forward modeling of M_i and that the overall functionality lies at the very primitive level, below the fully-fledged {semantics, syntax, phonology} configuration. Self-organization may be regarded as "intrinsic," as proposed by Triesch (2013).

be reached or when, inversely, the factor is too high and the responses have a hard time stabilizing, because overwhelmingly many new responses are constantly appearing. A teacher could elicit the proper responses to be learnt, however, by using specialized and/or personalized schemes that teach progressively intermediate responses that are closer to the spontaneous response of the subject or with the aid of special social interaction schemes, such as games between "fast" and "slow" learners, where the former act spontaneously as additional teachers to the latter. A series of simple computational experiments have verified these predictions [Tzafestas, 2008, Tzafestas, (in preparation)]. Our predictions and results indicate that generative, non-goal-directed, proto-imitation may contribute to a number of phenomena involving successful social learning, as regularly expected, or social learning deficits, such as isolation of a slow learner that cannot learn all the necessary responses. In the context of communication, such a mechanism may precede the emergence of actual communicative function, rather than communication being the end to which imitation is the means.

ASSOCIATION AND FUNCTION

Next, the passage to ***true*** meaning association can be envisaged where originally meaningless matching responses are associated to external objects and dynamic associations are built and maintained (Catmur et al., 2009; Waxman and Gelman, 2009; García et al., 2014; Heyes, 2015[2]; Sturdy and Nicoladis, 2017; Catmur and Heyes, 2019). The external references need not be pointed at or directly taught, although this might happen and can speed up learning (Eckerman and Stein, 1990; Ingersoll, 2010). Still, associations can be built spontaneously and reinforced in presence of multiple references in the environment, because statistically the correct reference for a given signal/response will be encountered much more frequently than other random references. We expect on average moderately rich environments in terms of wealth of stimuli to facilitate learning but cluttered ones to overstimulate and act as obstacles. Another prediction that was verified computationally (cf. above) is the emergence of multiple associations, in the sense of many responses associated with the same external reference, hence the basis for bilingualism and multilingualism. There are a number of additional intricacies concerning association. First, association has to be two-ways, from responses to references and from references to responses. Such associations can be developed either intrinsically in a Hebbian way and/or with the help of a reward-like contingency mechanism (Heyes, 2012; Cooper et al., 2013). In any case, how this could be implemented neurally would need to be worked out. Second, we need to carefully think the volume of associations that can be made. Birds, for example, can make a rather small number of associations, while the repertoire of a healthy human is comparatively enormous. Is big capacity a prerequisite for successful communication (for example, the capacity of the human vocal tract is indeed very big) or is it a drawback (because exploring a larger domain is harder)? It is also thinkable that reward/contingency mechanisms could be in the end necessary to learn a large set of responses or more complex responses, while purely Hebbian mechanisms could work for smaller sets or simpler responses.

OTHER PREDICTED BEHAVIORAL CONSEQUENCES

We can study accordingly extremes and deficits taken with the meaning/function association mechanism. First, reference salience is expected to be inversely proportional to the number of objects perceived in the environment and this can have drastic effects since extremely low salience would negatively affect the speed and ease of association and even block it altogether. Put otherwise, a subject that perceives too many things in the environment will be constantly distracted and hence slow or unable to learn, at least without special and personalized teaching. Such over-stimulation is thought sometimes to be the case in the autistic spectrum deficits (Remington and Fairney, 2017). Other predictions can be made as well. Bilinguals will be slower in learning concurrently their two first languages, but faster to learn the third. Because they are consistently stimulated twice as much and in a more complex manner than monolinguals, salience deficits will have less impact. Complete language replacement will be also hard, especially in the case of subjects with limited communicative repertoires.

All those predictions suggest that communicative and association deficits may sometimes arise even though the underlying imitation mechanism remains intact. We reiterate the case of the autistic spectrum disorders where social and communication deficits do not always go hand in hand with imitation problems and abnormalities (Leighton et al., 2008).

DISCUSSION

A number of assorted remarks can be made related to these ideas. First, association could be insufficient, but is it necessary in the first place? How else could meaning or function be assigned to external references provided the raw response imitation level exists? The alternative could be any top-down mechanism (Meltzoff and Prinz, 2002), such as innately given meanings waiting to be assigned (although again, some sort of limited association-like mechanism should be present for that) or an innate modular structure, refined at every step. We cannot rule out the possibility that some instances of such mechanisms exist in the human brain, however, as many authors argue, we think it is unlikely that they make the rule (Catmur et al., 2009; Jones, 2009; Froese and Leavens, 2014). Second, because we are interested in the behavioral predictions of our view, we are bypassing the sensorimotor

---

[2] Heyes (2001, 2012, 2015) and Cooper et al. (2013) insist on the non-goal-directedness of imitation and on the sufficiency of fundamental association mechanisms (more intricate than Hebb-like) to solve the correspondence problem (Brass and Heyes, 2005), i.e., the matching of (sensory) stimulus to (motor) response. This is what we think that happens at the level of our response imitation level. Hence at all levels there are associative processes continuously at play, and the lower level processes enable the higher level ones that only receive consistent input when the lower levels have achieved relatively stabilized dynamics.

correspondence problem and we assume that it is solved at the response imitation level in a basic associationist way. This would be an additional indication that selective mechanisms and especially association mechanisms appear at many places in the neural hierarchy.

Overall, we claim that a selective generative response imitation mechanism coupled with a higher level response association mechanism is capable of predicting many of the behavioral phenomena related to imitation in general, and especially a lot of abnormalities and deficits encountered in humans.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Ansermin, E., Mostafaoui, G., Beaussé, N., and Gaussier, P. (2016). "Learning to synchronously imitate gestures using entrainment effect," in *SAB 2016, LNAI 9825*, eds E. Tuci, A. Giagkos, M. Wilson, and J. Hallam (Aberystwyth: Springer LNAI Series). 219–231.

Brass, M., and Heyes, C. (2005). Imitation: is cognitive neuroscience solving the correspondence problem?. *Trends Cogn. Sci.* 9, 89–95. doi: 10.1016/j.tics.2005.08.007

Buzsáki, G. (2006). *Rhythms of the Brain*. New York, NY: Oxford University Press.

Catmur, C., and Heyes, C. (2019). Mirroring 'meaningful' actions: sensorimotor learning modulates imitation of goal-directed actions. *Q. J. Exp. Psychol.* 72, 322–334. doi: 10.1080/17470218.2017.1344257

Catmur, C., Walsh, V., and Heyes, C. (2009). Associative sequence learning: the role of experience in the development of imitation and the mirror system. *Philos. Trans. R. Soc. London B* 364, 2369–2380. doi: 10.1098/rstb.2009.0048

Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge, MA: MIT/Bradford Press.

Claidière, N., and Sperber, D. (2010). Imitation explains the propagation, not the stability of animal culture. *Philos. Trans. R. Soc. London B* 277, 651–659. doi: 10.1098/rspb.2009.1615

Cooper, R. P., Cook, R., Dickinson, A., and Heyes, C. M. (2013). Associative (not Hebbian) learning and the mirror neuron system. *Neurosci. Lett.* 540, 28–36. doi: 10.1016/j.neulet.2012.10.002

Eckerman, C. O., and Stein, M. R. (1990). How imitation begets imitation and toddlers' generation of games. *Dev. Psychol.* 26, 370–378. doi: 10.1037/0012-1649.26.3.370

Edelman, G. M. (1987). *Neural Darwinism. The Theory of Neuronal Group Selection.* New York, NY: Basic Books.

Froese, T., and Leavens, D. A. (2014). The direct perception hypothesis: perceiving the intention of another's action hinders its precise imitation. *Front. Psychol.* 5:65. doi: 10.3389/fpsyg.2014.00065

García, R. R., Zamorano, F., and Aboitiz, F. (2014). From imitation to meaning: circuit plasticity and the acquisition of a conventionalized semantics. *Front. Hum. Neurosci.* 8:605. doi: 10.3389/fnhum.2014.00605

Gergely, G., Bekkering, H., and Király, I. (2002). Rational imitation in preverbal infants. *Nature* 415:755. doi: 10.1038/415755a

Gerrans, P. S. (2013). Imitation, mind reading, and social learning. *Biol. Theory* 8, 20–27. doi: 10.1007/s13752-013-0112-4

Heyes, C. (2011). Automatic imitation. *Psychol. Bull.* 137, 463–483. doi: 10.1037/a0022288

Heyes, C. (2012). Simple minds: a qualified defence of associative learning. *Philos. Trans. R. Soc. London B* 367, 2695–2703. doi: 10.1098/rstb.2012.0217

Heyes, C. (2015). Homo imitans? Seven reasons why imitation couldn't possibly be associative. *Philos. Trans. R. Soc. London B* 371:20150069. doi: 10.1098/rstb.2015.0069

Heyes, C., Huber, L., Gergely, G., and Brass, M. (eds.). (2009). Evolution, development and intetional control of imitation. *Philos. Trans. R. Soc. London B* 364:1528. doi: 10.1098/rstb.2009.0049

Heyes, C. (2001). Causes and consequences of imitation. *Trends Cogn. Sci.* 5, 253–261. doi: 10.1016/S1364-6613(00)01661-2

Hodges, B. H. (2014). Rethinking conformity and imitation: divergence, convergence, and social understanding. *Front. Psychol.* 5:726. doi: 10.3389/fpsyg.2014.00726

Hurley, S., and Chater, N. (2005). *Perspectives on Imitation in Animals and Humans*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/5330.001.0001

Ingersoll, B. (2010). Broader autism phenotype and nonverbal sensitivity: evidence for an association in the general population. *J. Autism Dev. Disord.* 40, 590–598. doi: 10.1007/s10803-009-0907-0

Jones, S. (2009). The development of imitation in infancy. *Philos. Trans. R. Soc. London B* 364, 2325–2335. doi: 10.1098/rstb.2009.0045

Leighton, J., Bird, G., Charman, T., and Heyes, C. (2008). Weak imitative performance is not due to a functional 'mirroring' deficit in adults with Autism Spectrum Disorders. *Neuropsychologia* 46, 1041–1049. doi: 10.1016/j.neuropsychologia.2007.11.013

Meltzoff, A. N., and Prinz, W. (2002). *The Imitative Mind: Development, Evolution and Brain Bases*. New York, NY: Cambridge University Press.

Piaget, J. (1947/2001). La psychologie de l'intelligence. *The Psychology of Intelligence* (London; New York, NY: Routledge).

Pickering, M. J., and Garrod, S. (2014). Self-, other-, and joint monitoring using forward models. *Front. Hum. Neurosci.* 8:132. doi: 10.3389/fnhum.2014.00132

Remington, A., and Fairney, J. (2017). A sound advantage: increased auditory capacity in autism. *Cognition* 166, 459–465. doi: 10.1016/j.cognition.2017.04.002

Shea, N. (2009). Imitation as an inheritance system. *Philos. Trans. R. Soc. London B* 364, 2429–2443. doi: 10.1098/rstb.2009.0061

Sturdy, C. B., and Nicoladis, E. (2017). How much of language acquisition does operant conditioning explain?. *Front. Psychol.* 8:1918. doi: 10.3389/fpsyg.2017.01918

Triesch, J. (2013). Imitation learning based on an intrinsic motivation mechanism for efficient coding. *Front. Psychol.* 4:800. doi: 10.3389/fpsyg.2013.00800

Tzafestas, E. S. (2008). "On modeling proto-imitation in a pre-associative Babel," in *SAB 2008, LNAI 5040,* eds M. Asada, J. Hallam, J.-A. Meyer, and J. Tani (Osaka: Springer LNAI Series). 477–487.

Tzafestas, E. S. (2022). *A Selective-Associative Model of Imitation and Its Behavioral Consequences.* Partially Presented at Socializing Sensorimotor Consequences (SocSMC) 2018 Summit "Social cognition in humans and robots." Hamburg, September 2018.

van Baaren, R., Janssen, L., Chartrand, T. L., and Dijksterhuis, A. (2009). Where is the love? The social aspects of mimicry. *Philos. Trans. R. Soc. London B* 364, 2381–2389. doi: 10.1098/rstb.2009.0057

Waxman, S. R., and Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. *Trends Cogn. Sci.* 13, 258–262. doi: 10.1016/j.tics.2009.03.006

Whiten, A., McGuigan, N., Marshall-Pescini, S., and Hopper, L. M. (2009). Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee aspects of mimicry. *Philos. Trans. R. Soc. London B* 364, 2417–2428. doi: 10.1098/rstb.2009.0069

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 01 September 2021
doi: 10.3389/fnhum.2021.560657

Distinguishing Self, Other, and Autonomy From Visual Feedback: A Combined Correlation and Acceleration Transfer Analysis

Berkay Demirel[1]*, Clément Moulin-Frier[2], Xerxes D. Arsiwalla[3,4], Paul F. M. J. Verschure[4,5,6] and Martí Sánchez-Fibla[1]*

[1] Artificial Intelligence and Machine Learning Group, Department of Information and Communications Technologies, Universitat Pompeu Fabra, Barcelona, Spain, [2] Flowers, Inria and Ensta, University of Bordeaux and Paris Tech, Paris, France, [3] Department of Information and Communications Technologies, Universitat Pompeu Fabra, Barcelona, Spain, [4] Laboratory of Synthetic, Perceptive, Emotive and Cognitive Systems, Institute for Bioengineering of Catalonia, Barcelona Institute of Science and Technology, Barcelona, Spain, [5] The Barcelona Institute of Science and Technology, Barcelona, Spain, [6] Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain

In cognitive science, Theory of Mind (ToM) is the mental faculty of assessing intentions and beliefs of others and requires, in part, to distinguish incoming sensorimotor (SM) signals and, accordingly, attribute these to either the self-model, the model of the other, or one pertaining to the external world, including inanimate objects. To gain an understanding of this mechanism, we perform a computational analysis of SM interactions in a dual-arm robotic setup. Our main contribution is that, under the common fate principle, a correlation analysis of the velocities of visual pivots is shown to be sufficient to characterize "the self" (including proximo-distal arm-joint dependencies) and to assess motor to sensory influences, and "the other" by computing clusters in the correlation dependency graph. A correlational analysis, however, is not sufficient to assess the non-symmetric/directed dependencies required to infer autonomy, the ability of entities to move by themselves. We subsequently validate 3 measures that can potentially quantify a metric for autonomy: Granger causality (GC), transfer entropy (TE), as well as a novel “Acceleration Transfer” (AT) measure, which is an instantaneous measure that computes the estimated instantaneous transfer of acceleration between visual features, from which one can compute a directed SM graph. Subsequently, autonomy is characterized by the sink nodes in this directed graph. This study results show that although TE can capture the directional dependencies, a rectified subtraction operation denoted, in this study, as AT is both sufficient and computationally cheaper.

Keywords: theory of mind, cognitive development, autonomy, attention, agency, sensorimotor learning, developmental psychology, computational cognition

Edited by:
Filippo Zappasodi,
University of Studies G. d’Annunzio Chieti and Pescara, Italy

Reviewed by:
Sebastiano Stramaglia,
University of Bari Aldo Moro, Italy
Joseph Thachil Francis,
University of Houston, United States

Correspondence:
Berkay Demirel
berkay.demirel@upf.edu
Martí Sánchez-Fibla
marti.sanchez@upf.edu

Specialty section:
This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience

Received: *09 May 2020*
Accepted: *14 July 2021*
Published: *01 September 2021*

Citation:
Demirel B, Moulin-Frier C, Arsiwalla XD, Verschure PFMJ and Sánchez-Fibla M (2021) Distinguishing Self, Other, and Autonomy From Visual Feedback: A Combined Correlation and Acceleration Transfer Analysis. Front. Hum. Neurosci. 15:560657. doi: 10.3389/fnhum.2021.560657

1. INTRODUCTION

We are just beginning to uncover the mysteries of how the brain makes sense of the surrounding world and learns to perform social interactions during development. Babies are very sensitive to motion and use it to organize visual scenes into higher-order structures and seem to rely on instantaneous, immediate motion co-occurrences (Gelman, 2003; Luo et al., 2009). Babies learn to recognize themselves and other agents (Gelman, 2003) and develop a notion of autonomous entities that can actively function by themselves (like persons or animals), or need to be actuated by others (like objects and toys). That is to say, babies learn to distinguish animate from inanimate entities early on Luo et al. (2009) and Opfer and Gelman (2011) (thus we can also say that babies learn to assess whether an entity is autonomous or not). In addition, there is evidence that the brains of vertebrates have ancient neural mechanisms susceptible to the detection of animacy, as they call it in Mascalzoni et al. (2010) study. Learning these distinctions is a basic prerequisite of how babies acquire social cognition (Baillargeon et al., 2015), which is closely tied to the acquisition of a Theory of Mind (ToM), that is, learning to assess and predict beliefs, intentions, and goals of others. Recent developments in AI (based on Bayesian probabilistic inference Baker et al., 2011 and Deep Learning Rabinowitz et al., 2018) have attempted to address the challenge of learning to acquire a ToM, that is assessing and predicting the beliefs, intentions and goals of other agents (refer to also Freire et al., 2018, 2019 for a control theoretic perspective). ToM can be assessed at many levels, but a reliable ultimate test should be able to answer questions about states of the environment of the agent (including other agents), as in Nematzadeh et al. (2018), in which neural models augmented with external memory structures are evaluated in question answering. Indeed, the underlying question is whether cognitive mechanisms underlying action and perception in the physical world can somehow be generalized to the social setting, involving beliefs and intentions of other agents (Arsiwalla et al., 2016, 2017a,b; Verschure, 2016). The long-term goal of this research agenda is to provide a computational basis for how ToM abilities could arise from low-level sensorimotor (SM) interactions, that is, bottom-up from agent interaction behaviors (Freire et al., 2018, 2019). In this sense, we differ from multi-agent approaches like (Marsella et al., 2004), where beliefs are symbols added as logical facts and inference is performed through rule systems. We also distinguish this minimal correlational/temporal low-level SM approach from learning complex parameterized neural nets like (Rabinowitz et al., 2018). Presumably, in a later phase of learning and development, and *via* building up from the bottom up interactions, ToM abilities may be refined *via* top-down optimization of social behavior and cooperation.

Acquiring ToM abilities requires labeling and clustering the SM data stream of the interaction of the agent in terms of which visual features belong to its own body, which ones belong to other entities, which ones can be controlled by its actuators, and which ones can be controlled by themselves or are passive and need others to move. The problem of deciphering self from others in robotics and AI has been addressed by several computational models in studies such as Brody et al. (2017), Thomas et al. (2017), Sánchez-Fibla et al. (2017b), Rybkin et al. (2018), and Pertsch et al. (2018). We approach the labeling problem from the perspective of identifying what are the minimal requirements to distinguish self, other, and autonomous or passive entities from visual feedback alone. Under the Gestalt principle of common fate: "what moves together, clusters together," we show that a simple correlational analysis of visual pivots' velocities, can suffice to distinguish "self" and "other" (refer to sections 2.4, 3.1). This happens in the long term, as in the short term, the different parts of the same body may not move together as they may be actuated by different joints moving in different directions.

We subsequently introduce an Acceleration Transfer (AT) measure (Sections 2.6, 3.2) targeting the extraction of directional dependencies between visual pivots for the detection of autonomy defined as the capacity of entities to move themselves. Pairwise directional dependencies cannot be captured by correlation analysis. Addressing the detection of autonomy or animacy (as named by Mascalzoni et al., 2010 and widely addressed by the Developmental Psychology literature Luo et al., 2009; Opfer and Gelman, 2011; Baillargeon et al., 2015) can be considered a novelty and a contribution from a computational modeling perspective (refer to section 2.7). To the best of our knowledge, we are not aware of any other research that state what are the minimal computational requirements for detecting autonomy, which we postulate could be the AT among visual motion pivots (our SIPs). Acceleration is directly linked to force (through Newton's 2nd law), and it reflects shorter events in time, compared to velocity. In addition, there are connections to studies providing neurophysiological evidence of acceleration responsive neurons as in Schlack et al. (2007).

The underlying principle of AT measure to detect autonomy is that inanimate entities can only receive acceleration and can never create it by themselves; so, passive entities can accelerate but only through others. It would seem like we are not strictly required to look at pairwise dependencies, but the acceleration of an inanimate object may have been caused by another entity several steps before; so, we need to assess the dependencies of all pairwise interactions. To make fewer assumptions, we face the autonomy identification problem without relying on haptics or filtering by the proximity of visual pivots. Thus, we switch to an alternative underlying principle: inanimate objects will always be a sink in the directional graph of pairwise interactions defined in section 2.8. The so-called, SM graph can be computed from the AT and other standard measures, such as Granger causality (GC) or transfer entropy, although we show that AT has some advantages as discussed in section 3.3. We evaluate and validate both the velocity correlation and AT analysis (with the estimation of directed SM graphs) in a bi-manual, multi-agent, the ball sliding task (freely available[1], refer to **Figure 1** and section 2.2 for details).

2. METHODS

We define the components that are going to constitute the sensory space and the motor actuators *via* a simulated environment that will emulate a newborn in the presence of passive objects and another agent. We will describe which sensory and motor signals we will consider and we will build a methodology to address how one can minimally extract, and make sense, the SM stream of data. The particularity of this method is that we address this problem solely from the visual modality. But we do not follow a typical deep learning approach (Mnih et al., 2015), instead, we identify the minimal elements

---

[1] The simulation can be downloaded from https://github.com/santmarti/PythonRobot2DSim. A video TwoArmSetup.mp4 is also available under the folder videos in the same github repository. Notebooks for the generation of figures of the paper will also be made available.

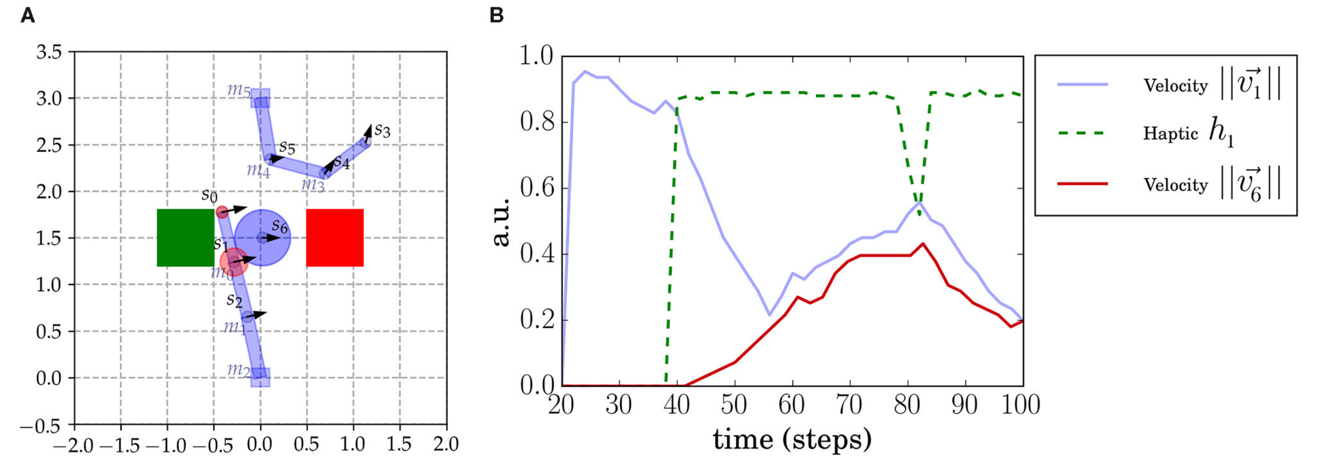


FIGURE 1 | Sally-Anne robot setup. **(A)** Two 3DOF robotic arms face each other and, in the middle, an object is constrained to move in a horizontal axis and cannot be seen when under the boxes (indicated by the green and red squares). The agents perceive the sensory interest points (SIPs) $s_0, ..., s_6$ placed in the arms and object. Agents also receive tactile feedback when contacts occur (indicated as red circles). We also annotate the different motors (m_0, m_1, m_2 for the lower arm and m_3, m_4, m_5 for the upper arm) placed in each join. The haptic signal of every SIP (that we denote h_i) is simulated and computed from proximity to the real contact point (which may be between SIPs). The h_i intensity is depicted with a red circle centered at the corresponding SIP position. **(B)** Sensorimotor time series in a moment of contact of SIPs s_1 and s_6. The haptic signal h_1 (green dashed line) increases suddenly at the moment of contact (at 40 time steps). At this point the velocity of the object increases as well.

from which we can start reasoning about visual motion cues, through visual pivots reminiscent of the biological motion (BM) pivot dots (which are in turn related to social cognition (Pavlova, 2011).

2.1. Notation

The visual modality is reduced to a set of relevant sensory interest points (SIPs), as we formalize in this study based on Sánchez-Fibla et al. (2019). Reducing the visual input to a set of SIPs does not seem to be a limiting assumption, as humans and animals, in general, have the ability to understand and recognize action from the observation of a few dots attached to relevant parts of a body in motion, also called BM. BM is also related to social cognition as patients with deficits in social interaction are also compromised in visual body motion processing (Pavlova, 2011).

A SIP is attached to a relevant feature of the visual field (i.e., the end point effector, at the junction between two arm links, at the corner of a square object, etc.). Refer for example the prototypical SIPs that we chose for this setup in **Figure 1**. SIPs can be computed from a stream of images *via* different methods (including deep learning techniques), but a method that would provide direct mapping is the so-called Scale Invariant Feature Transform (SIFT) points, which provides local feature pivot points (Lowe, 1999).

A sensory state s consists of a set of SIPs: $s = \{s_0, \cdots, s_i, \cdots\}$ each s_i being $\{p_i = \{x_i, y_i\}, h_i\}$ where p_i corresponds to its 2D Cartesian coordinates and h_i a real value haptic signal normalized from 0 to 1. The changing signal information for each SIP can be extracted from the temporal SM data: $\Delta s_i = \{\vec{v_i} = \{v_i^x, v_i^y\}, \vec{a_i} = \{a_i^x, a_i^y\}, \Delta h_i\}$

Where $\vec{v_i}$ and $\vec{a_i}$ denotes the velocity and acceleration vector of s_i, respectively. We are going to assume that vectors are normalized, $||\vec{v_i}|| \in [0..1]$, $||\vec{a_i}|| \in [0..1]$. Δh_i is the haptic signal change. In continuation we will consider and refer to time series of positional information of SIPS (x, y coordinates), velocities of SIPs ($\vec{v}$ velocity vector with v^x, v^y components and magnitude $||\vec{v}||$), and accelerations ($\vec{a}$ acceleration vector with a^x, a^y components and magnitude $||\vec{a}||$).

The motor apparatus of a newborn includes over 650 skeletal muscle actuators. We restrict this study to a small number of actuators. A motor state $m = \{m_0, ..., m_{n-1}\}$ corresponds to the angles of every joint $m_i \in [0...2\pi]$. The motor space is denoted by $\mathcal{M}$. A delta motor state $\Delta m = \{\Delta m_0, ..., \Delta m_{n-1}\}$ is the velocities of each joint, n being the total number of DOFs.

2.2. Implementation of the Sally-Anne Paradigm

The so-called Sally-Anne setup is a robotics implementation (depicted in **Figure 1A**) inspired from the psychological experiment with the same name, designed to probe attribution of beliefs and was first studied in relation to autism (Baron-Cohen et al., 1985) (although the result stating that autistic subjects fail at the "Sally-Anne test" is still under debate Tager-Flusberg, 2007). Let us briefly summarize the experiment. The participant is presented with two boxes and an object and an imaginary character, Sally, places the object in box A and leaves the room. Another imaginary character, Anne, puts the object in the other box B. Sally comes back and the participant is asked, in which box will Sally look for the object? The test is passed if box A is chosen, as, although we know that the object is in box B, Sally cannot know it because she did not see Anne transferring the object to the other box. This paradigm thus probes the ability of the participant to model belief states of the other.

This setup here, inspired by the Sally-Anne experiment, is a synthetic computational implementation of the Sally-Anne paradigm. It consists of two agents (two robot arms), an object and two colored boxes or areas that only occlude the entities underneath (refer to **Figure 1A**). The two-arm robots face each other so that there is a region in the space where both interact, thereby affecting each other's actions. Depending on the current state, each agent can move freely, interact with the object alone, interact with the other agent, or interact with the other agent using the object. The object is a sliding ball that is constrained to move in an horizontal line between the two boxes. Each entity in the setup has associated sensory points (attached to joints in the robot arms and the object). In total, seven SIPs are considered: three in each arm of the robots and one in the object. End effector SIPs are denoted by $s_{left} = s_0$ and $s_{right} = s_3$. A SIP is added at the object: $s_{obj} = s_6$ at the center. Each agent perceives the totality of the visual cues of the scene (all SIPs, from s_0 to s_6, the object) and their haptic signals. Each robot arm has three joints that can be actuated independently by each agent.

The haptic signals in this setup are simulated and are computed considering their distance to the closest contact point (in **Figure 1A** haptic signals are represented by red circles centered at its SIP). The haptic signal h_0 seems to increase inconsistently, but, as shown in **Figure 1A**, h_0 is faint on initiation of contact and increases slowly because it gets closer to the contact point between the two shapes (rectangle and circle). On the other hand, h_1 is closer to the real contact point so it increases rapidly, as shown in **Figure 1B**.

2.3. Data Generation

Sensorimotor data is generated from a mixture of behaviors: motor babbling and goal-directed movements targeting the object. Motor babbling assigns random velocities to the motor joints of the arms (Δm_i for $i \in 0, 1, 2$ for the bottom agent and $i \in 3, 4, 5$ for the top one) at different moments. We need independent motor babbling to not cause an artificial correlation between the agents due to the synchronous randomly generated behavior. We sustain, for different time steps, the assigned motor torque to each joint m_i. We need to sustain the torque applied to generate a consistent movement with a consistent effect during a certain time. Because of the characteristics of the setup, we needed to set certain rules on top of the independent random signals sent to the motor joints to generate a rich and representative dataset. For example, when the object reached an end-point under one of the two boxes, the object was reset in the middle again, and the two arms were positioned at their initial positions. As an example, in **Figure 1B**, we show the time series of s_1 and s_6 SIPs before and during an object contact.

Generating the behavior of agents in this way, we acquire a dataset of all SIP data streams (position, velocities, accelerations, and haptic contacts) and we are thus ready to apply the corresponding statistical tools to determine the nature of every SIP.

2.4. Correlational Analysis

We start by presenting a minimal method to assess which elements of the visual sensory scene correspond to the self and the rest (other entities, objects, or other agents). We do so by a correlational analysis of the different streams of data from the generated dataset. The Pearson's correlation (or simply correlation) coefficient accounts for the linear relationship between a set of points with (x, y) coordinates. In general, these coordinates are the outcome of two random variables. In this case, we look at the correlation between SIP measures, like x coordinates, velocities, accelerations, and we also consider their haptic feedback values. The correlation between two random variables X and Y is then defined by:

$$\rho_{X,Y} = \mathrm{corr}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\mathbb{E}[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

where μ_X, μ_Y are the respective variable means and σ_X, σ_Y are the SD or variances. Correlation is undefined when either one of the variances is 0 (division by zero). This happens, for example, when one variable is a constant value. This can happen in this case when looking for correlations of the end-point of an arm s_0 and the object s_6, in the case where the object remains static and untouched.

We aim to stick with the simplest methods possible to extract self/other/object characteristics from SM data. But, of course, correlation comes with its limitations. First, Pearson's correlation captures the amount of linear relationship between two variables and can have some trouble in detecting dependencies that are not strictly linear, like the ones between motor signals m_i and sensory point coordinates of SIPs s_i. Second, correlation does not allow us to infer temporal or causal relationships between sensory events. A correlational analysis does not provide directional information as the Pearson correlation coefficient is symmetric: $\mathrm{corr}(X, Y) = \mathrm{corr}(Y, X)$.

2.5. Temporal Dependencies

In the context of correlational analysis, the one obtained by computing $\mathrm{corr}(X, Y)$, from data time series extracted from SIPs, the temporal structure of data is completely discarded. The sampled values of variables X and Y have no temporal relationship and they are only inspected in pairs at the same time point t. There are many ways of considering temporal (and order) information. The main principle used when doing this is to assess what a second-time series adds to the prediction of another one. What is the added prediction value of time series $x^{0..n}$ to time series $y^{0..n}$? Consider the sample at time t, y^t. How the samples x^{t-i} help to predict y^t in addition to y^{t-i}.

Let $x^{0...n}$ and $y^{0...n}$ be stationary time series. $x^{0...n}$ is said to Granger cause $y^{0...n}$ if it has an added value in the prediction of $y^{0...n}$. If values of $x^{0...t-1}$ add explanatory power to y^t (in addition to $y^{0...t-1}$), $x^{0...n}$ is said to Granger cause $y^{0...n}$.

Granger causality is based on a linear correlation test between past and current values, implemented with a statistical t-test. We based all our previous analysis on correlations, which only capture linear relationships, and these were sufficient to extract the proximo-distal and motor to sensory relations. It seems a perfectly logical approach to use GC to capture directional relationships.

Transfer Entropy (Vicente et al., 2011) can deal with non-linear relationships, and many indicate that it is more suitable for quantifying causality (Razak and Jensen, 2014). TE comes with its drawbacks: taking into account just a small-time history, necessary for its computation, it becomes inefficient to compute and demands a lot of data to have maximum exposure to all history combinations.

Both measures, GC and TE, have an additional parameter which is the lag or history time steps that the computation takes into consideration. The lag is an additional parameter that in the case of TE is problematic as it demands an increasing exponential number of samples to be accurate (as the history increases).

Granger causality and TE are based on distribution of the sampled data, and we need a measure to operate on a trial-by-trial basis. Thus, we cannot ensure stationarity of the time series when dealing with isolated trials. Refer to **Figure 5** for example, in which we show a single approach and contact trial of one arm and the object. Although we can make linear piece-wise relations (using the modulus/magnitude of velocity and acceleration vectors), the effect detection mechanism that we are looking for is not linear as a whole. When two SIPs enter in contact, there is first an anti-coupling followed by a progressive coupling after contact (we explain this in more detail in continuation).

2.6. Acceleration Transfer Measure

To overcome some of the drawbacks of GC and TE (as discussed in previous section 2.5) which also hold for other measures, such as Copula-Based measure (Junker et al., 2019), we define an alternative instantaneous AT measure. By instantaneous and not temporally dependant we mean that the AT measure does not have a lag/history parameter because it can be computed at a single time step. The AT measure is computed according to an instantaneous positive subtraction of accelerations between two SIPs. Its computation follows the intuition that if a SIP is losing acceleration it can be because it is being transferred to another SIP. As Newton's law states ($F = m.a$), acceleration is proportional to force, meaning that AT can be seen as a form of force transferred or being applied to.

Consider $a_i^{0...t...n}, a_j^{0...t...n}$ to be the sampled accelerations of SIPs i and j from time step 0 to n. Then, the AT transfer from SIP i to j at time t will simply be the subtraction of accelerations at time t: $T(a_i^t, a_j^t) = a_j^t - a_i^t$. We could then consider the AT between the two SIPs $\mathcal{A}_{v0}(a_i, a_j)$ to be the sum over all the time steps $\mathcal{A}_{v0}(a_i, a_j) = \sum_{k=0}^{t} T(a_i^k, a_j^k)$. We are going to consider $\mathcal{A}_{v0}$ to be the simplest version 0 of the measure. We are then going to add three filters that will target the AT events that we want and will constitute versions 1, 2, and 3. The final AT measure will contain all filters and will be considered version 4.

- Version 1. We are going to filter the instantaneous transfer by the minimum of the two signal values, that is, the transfer cannot be bigger than the minimum of the two signals. The intuition behind this filter is that if signals are very different in magnitude we should be careful not to allow a too big transfer also because the signals may not be related.

$$T(a_i^k, a_j^k) = \begin{cases} min\{a_j^k - a_i^k, min\{|a_i^k|, |a_j^k|\}\} & \text{if } a_j^k - a_i^k > 0 \\ max\{a_j^k - a_i^k, -min\{|a_i^k|, |a_j^k|\}\} & \text{otherwise} \end{cases}$$

- Version 2. If one signal is close or equal to 0, then, the instantaneous transfer is also set to 0. This filter can be enforced by adding a multiplying factor to the instantaneous transfer $T(a_i^k, a_j^k) = (a_j^k - a_i^k) * min\{1, |a_i^k||a_j^k|\}$
- Version 3. Signals should have opposite signs. We consider only transfers in which one signal is losing acceleration while the other is gaining it. This is because we are interested in collisions and moments of energy exchange. We enforce this filter with the following inequality that needs a small epsilon (i.e., $\epsilon = 0.01$) to filter out signals of the same sign:

$$T(a_i^k, a_j^k) = \begin{cases} a_j^k - a_i^k & \text{if } |a_i^k| + |a_j^k| \geq |a_i^k + a_j^k| + \epsilon \\ 0 & \text{otherwise} \end{cases}$$

The AT measure that we use (version 4) includes all filters. We analyze the effects of the three filters in the Results section 3.2 applied to different synthetic generated signals.

The AT measure can be refined further using filters of visual proximity or haptic signals from SIPs h_i, with the additional assumptions that there can not be a transfer of acceleration between two SIPs that are not touching or that are not in close visual proximity of each other. We prefer to address the more general setting by limiting the number of underlying assumptions, and focusing on showing the viability of this measure as a computationally cheaper and sufficient alternative rather than a specific implementation. Haptic signals as well as visual proximity are nevertheless strong canditates for additional assumptions in developmental settings and can be the basis of further research.

2.7. Detecting Autonomy

We define the autonomy (of an entity) as the ability to move by itself. In the setup that we present, two autonomous entities constitute the lower (with associated SIPs s_0, s_1, s_2) and upper (associated SIPs s_3, s_4, s_5) arms and both are controlled by independent motor signals (m_0, m_1, m_2, and m_3, m_4, m_5 respectively). The object and its associated SIP s_6 is a passive entity, non-autonomous, it cannot move by itself but at certain moments it can appear to be moving without any contact (in a low friction environment for example). Thus, the main characterizing property of a passive entity is that it never starts moving without a previous contact, in other words, it never accelerates by itself: an active entity must inject kinetic energy into it. To assess autonomy we need to go beyond correlation and distinguish which SIPs would be dead-ends in the path of directed relations.

2.8. The Sensorimotor Graph

Sensorimotor graphs can be constructed in different ways. For instance, in the context of the Distributed Adaptive Control (DAC) (Duff et al., 2011), a directed-graph network stores a compacted version of past sequences of experiences. In the graph, nodes are SM couplets including the reward of the experience of the agent and edges are reinforced when experiences have co-occurred. Other approaches consider SM graphs as memory structures: Toussaint (2006) presents a similar SM directed-graph network approach.

In this study, we consider a SM graph of a different kind. The visual features accessed through the so-called Sensory Interest Point data (SIP coordinates, velocities, accelerations, haptic signals) are considered random variables, and the graph corresponds to the dependencies between these variables. Nodes of the graph are SIP-related variables and edges are extracted dependencies between them (correlations for instance). Indeed, the matrix of pairwise correlations can be binarized using a threshold, thus obtaining an adjacency matrix defining an undirected graph as correlation is symmetric. Using AT, we can produce an asymmetric matrix thus leading to a directed graph. The same is true for any other measure like GC and TE. From their matrices of pairwise dependencies (that we explain in section 3.3), one can extract a directed SM graph.

The directed SM graph can be the fingerprint of autonomy as an inanimate entity should not be capable of transferring acceleration to other entities; thus a non-Autonomous entity will always appear as the sink of a directed SM graph. An inanimate/non autonomous/passive entity is characterized as being the end node in the directed SM graph of the SIP AT measure. There might be rare occasions where this assumption is violated, such as in the case of a billiard ball, where we transfer acceleration to the white ball, which in turn transfers acceleration to other balls, but always losing energy, never gaining it.

3. RESULTS

In the current section we present numerical results obtained from the SM data set that we extracted from the Sally and Anne setup (explained in section 2.3). We first present a descriptive interpretation of the correlation analysis of the Sensory Interest Point (SIP) interactions (refer to section 3.1 as seen previously in Sánchez-Fibla et al. (2017b). From the correlation limitations described, we provide results (section 3.2) of the newly introduced AT measure (Methods section 2.6) along with a comparison of the measure with GC and TE measures. The SM Graphs are alternative ways of visualizing the pairwise interactions between SIPs according to the different measures considered.

3.1. Correlational Analysis

We found that patterns of correlations arise between velocities of SIPs (**Figure 2A).** A signature of the proximo-distal organization of the arm of each agent is present. The closer the two joints are, the greater their correlation is: corr(v_0^x, v_1^x) is less than corr(v_1^x, v_2^x) for the bottom agent and corr(v_3^x, v_4^x) is less than corr(v_4^x, v_5^x). This inter agent correlation (proximo-distal) pattern is nearly identical between the two agents. Both include the same chessboard pattern as observed in **Figure 2A** top-left and bottom-right (excluding v_6 row and column which corresponds to the SIP in the object). By matching their own joint velocity correlation pattern with the one observed from another agent, this could provide a first level of mirroring between agents, where each one is able to match its kinematic structure with that of the other.

We also observed a correlation of the object velocity v_6^x with both arms having different intensities following the same proximo-distal pattern under different conditions: (i) when not in contact (**Figure 2B**), (ii) when in contact with the lower arm (**Figure 2C**), (iii) when in contact with the upper arm (**Figure 2D**), and (iv) when in contact with either one arm or the other (**Figure 2E**). This pattern may also provide a second level of mirroring, where agents can discover that similar movements of the other provide similar effects on the objects, paving the way to a notion of shared affordances, useful for joint action planning.

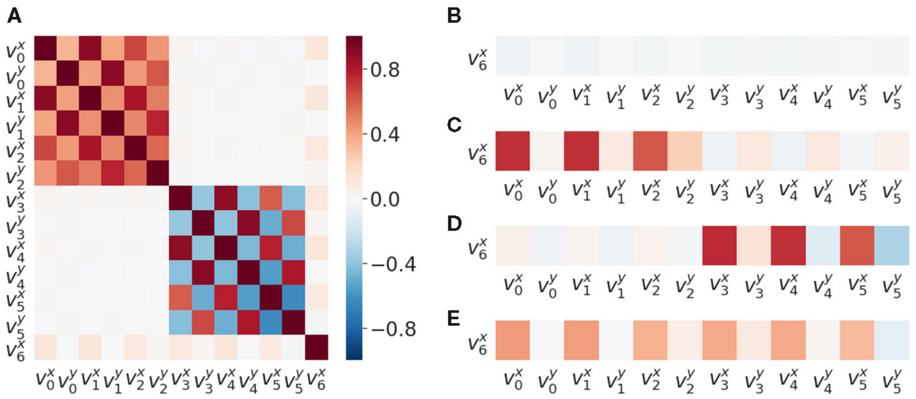


FIGURE 2 | Velocity correlation matrices of SIPs.. **(A)** Velocity correlation matrix of all the SIPs generated from 50,000 random SM interactions. **(B)** Correlation between SIP velocities, object velocity v_6^x and arms $v_0^x, ..., v_5^x$, when there is no contact (no haptic signal active). **(C)** Correlation between object and arms SIP velocities when there is contact between the lower arm and the object. **(D)** Correlation between object and arms SIP velocities when there is contact between the upper arm and the object. **(E)** Correlation object and arms SIP velocities when there is contact between the object and either arm.

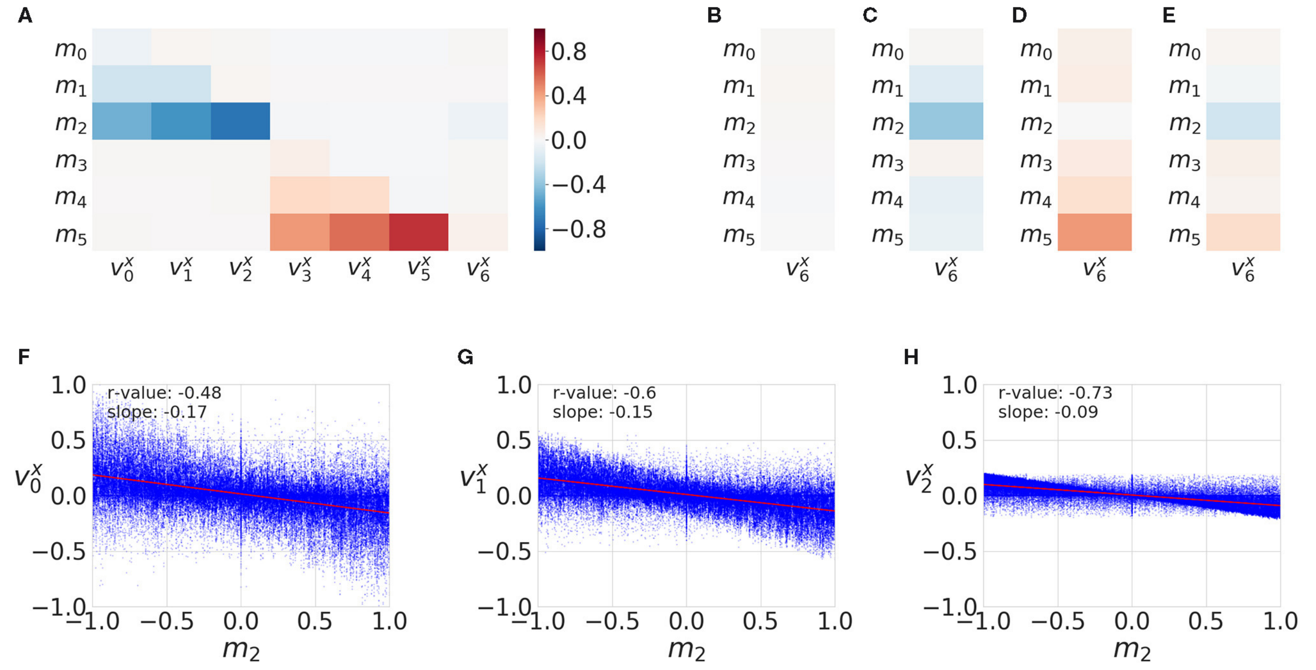


FIGURE 3 | Motor and sensory correlations. **(A)** Correlations matrix between motor activation (m_0...m_5 signals) and SIP velocities v_0, ...,v_6, generated from 50,000 random SM interactions. **(B)** Correlation column of all motor activations with the object SIP velocity v_6^x for SM data when there is no contact (no haptic signals active) **(C)** Correlation of motors with object SIP velocity when there is contact between the lower (bottom) arm and the object. **(D)** Correlation of motors with object SIP velocity when there is contact between the upper arm and the object. **(E)** Correlation between motor and object SIP when there is contact between the object and either arm. **(F–H)** All data points and linear fit for *m*2 and SIP velocities v_0^x, v_1^x, v_2^x. The *r*-values correspond to the equivalent correlation matrix color values.

We found motor signals to be correlated with velocities of SIPs (**Figure 3A**). The bottom joint of the bottom agent (m_2) is strongly negatively correlated with the x velocity components of sensory points s_0, s_1, and s_2, because this joint moves the whole arm. The same happens with the top agent. This matrix can assess controllability characteristics and potentially restricts the forward model to be learned for the relevant SIP signals: i.e., the agents could filter out SIPs that are not correlated to its available motor signals.

The object x velocity component v_6^x does not appear to correlate with any motor signals (**Figure 3A**) because, in this study, we consider the totality of the interaction (with and without contact). For this reason, in **Figure 2** we distinguish the different conditions with regard to haptic signals: (i) when no contact is made (**Figure 3B**), (ii) when contact with the lower arm is made (**Figure 3C**), (iii) when contact with the upper arm is made (**Figure 3D**), and (iv) when contact with either arm is made (**Figure 3E**). This motor to object correlation (when in contact with each arm) is reminiscent of the notion of affordance (Sánchez-Fibla et al., 2011), as it characterizes the effect of a motor action into a movement characteristic of the object.

In addition, the correlations change sign in the x coordinate for the top and bottom arms. This is because the arms are initialized to different positions and one has a tendency to touch the object in one direction and the other in the opposite one.

Summarizing the correlational velocity analysis captures well the principle of ‘"what moves together clusters together” on the long term and is capable of distinguishing self/other (proximo-distal relations) and motor to sensory dependencies.

3.2. Acceleration Transfer

We present results of the AT measure introduced in section 2.6, which computes an integration in time of instantaneous acceleration transfers, that is a subtraction of accelerations at a given time step plus a series of filters explained in section 2.6. We plot in **Figure 4** an evaluation of all versions of the AT measure (columns) for different generated signals (rows), which all include a small normally distributed noise. In the first row, we plot two signals that include a Gaussian of opposite signs. The physical interpretation of these signals could be the end-point of an arm making contact with the object. The object would then gain acceleration (y signal in the figure) and the end-point of the arm would lose it (x signal). These signals can be considered a crude approximation of AT from an animate object to an inanimate object (refer to **Figure 5** explained below). In the second row, we add some lag to one of the signals emulating some sort of compliance or delayed actuation. We observe that the transfer starts to fade away with a greater lag. In the third row, we plot two Gaussians of the same sign. A possible physical interpretation is that two SIPs are hit by a third one providing energy to both. It makes sense to filter out this situation as it does not correspond to an exchange of acceleration (done by the filter introduced in Version 3). The fourth row shows a variation of the previous situation where there is a lag between the two same signed signals. Fifth row depicts two random signals of the same sign. The last row of **Figure 4** reveals a weakness of the AT measure as it may detect continuous false-positive transfers from random signals of opposite signs. These random oscillations are not typical of the minimal jerk movement trajectories of

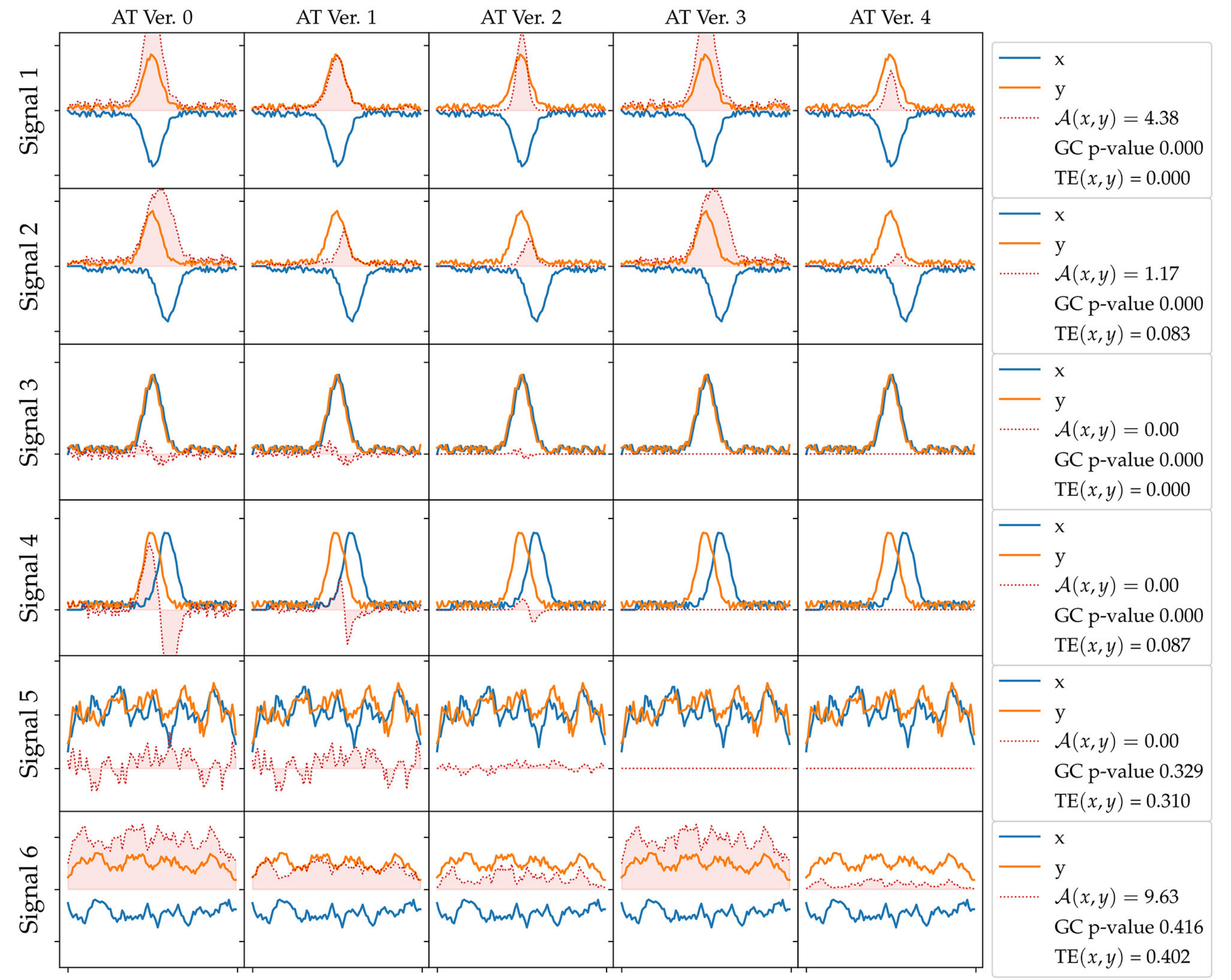


FIGURE 4 | Acceleration Transfer (AT) measure versions evaluated on different synthetic generated signals. The rows correspond to different generated signals: 1) two Gaussians of opposite signs, 2) two Gaussians of opposite signs with a lag, 3) two Gaussians of the same sign, 4) two Gaussians of the same sign with a lag, 5) random smoothed signals, 6) random smoothed signals of opposite signs. The columns correspond to the different versions of the AT (refer to text for further details): V0) signal subtraction, V1) filter by minimum value, V2) filter by closeness to 0, V3) filter by value sign, and V4) all Filters included. In the legend at each row, we show the AT value, the Granger Causality (GC) *p*-value, and transfer entropy (TE) for each signal. A GC $p < 0.05$ means *x* signal Granger causes *y*. TE is in bits.

human behavior or the ones that one can program in a robot arm.

We continue with a prototypical example of an arm approaching and making contact with the object. **Figure 5** shows an example of the AT measure and how it captures the transfer of acceleration from the lower arm (SIP s_0) to the object (SIP s_6). At the moment of contact, SIPs from the arm (acceleration time series a_0) and the object (acceleration a_6) anti-correlate, giving rise to the moment of acceleration transfer, and then synchronize just afterward, both tending to 0 after the force of the initial contact is exhausted and they continue to move at a constant velocity. AT is 0 in this last part because of the filter applied in version 3 that discards transfers of the same sign. AT is 0 before contact and at the end because of the filter applied at version 2 that sets to 0 the transfer if one of the two signals is close to 0. The different version filters of the AT measure are explained in section 2.6.

We also computed the pairwise AT measure (defined in section 2.6) between all acceleration a_i signals. The results are shown as a matrix in **Figure 6A**. The matrix is computed from 200 repetitions of a random moving arm (with sustained movements for a random number of steps) and the other arm performing a goal-directed approach to the object. Each trial, the arm that moves is chosen randomly with a 0.5 probability. The AT measure with its filters (refer to section 2.6) becomes very specific, and it is able to catch the events we are interested in, which are the object interactions. AT does not detect any interactions between the arms SIPs.

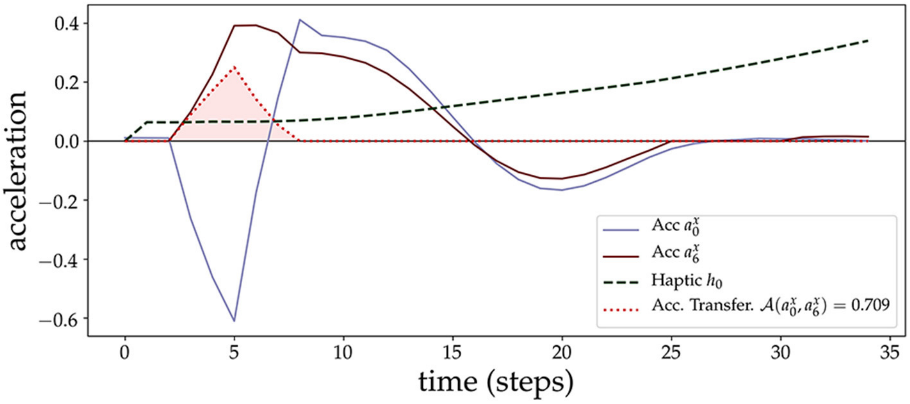


FIGURE 5 | Acceleration Transfer measure. We plot an example of the AT measure (shaded area in red) due to a contact of the lower arm (sensory point s_0, with corresponding acceleration a_0^x) with the object (sensory point s_6, with corresponding acceleration a_6^x). At the beginning of the contact, the arm loses acceleration rapidly which is transferred to the object. After some time steps (after time step 10), both start to couple and synchronize.

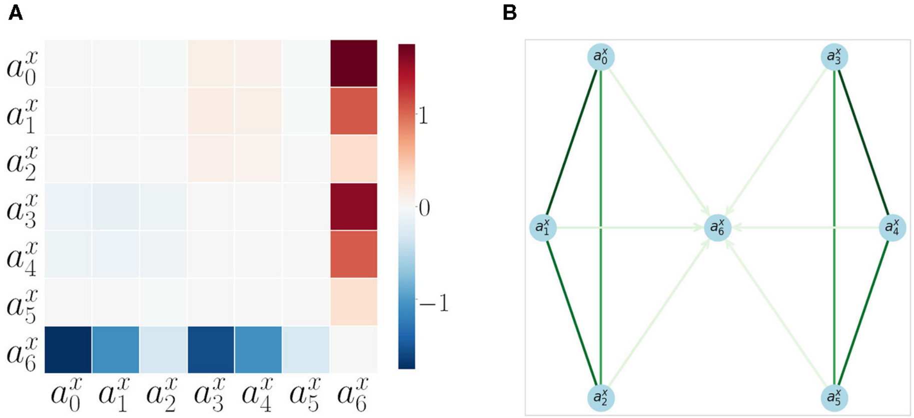


FIGURE 6 | Acceleration Transfer results **(A)** Matrix of pairwise AT measure values between all SIPs for all SM interaction data. **(B)** Directed graph of SIP accelerations a_0^x, ...,a_6^x, generated from 200 repetitions of sensorimotor interactions. The edge colors show the correlation strength between the nodes (extracted from the velocity correlation analysis, refer to section 3.1), while the arrows show the direction of AT (refer to section 3.2). Only arrows with a significant AT are plotted. Take note that the node a_6^x is the only terminal node, as it can not accelerate by itself and depends on the AT from other SIPs.

When the AT is applied to the SM data, that is gathered with motor babbling, it usually happens that two SIPs accelerate and decelerate synchronously, and, thus, AT detects lots of false positives. In addition, if random movements are not sustained in time (jerky movements), the AT measure detects lots of false positive transfers as predicted by the random synthetic signals analysis of the last row in **Figure 4**.

The results are also shown in **Figure 6B** as a directed graph: the strength of the arrows comes from the previous correlation results (as AT detects no specificity between arms). The direction of the arrows comes from the AT measure. No arrows are plotted for inter-arm relations as AT is not able to capture them. The object appears as the end node (the sink) of the directed graph, proving that it always receives acceleration and never produces it: the signature for a measure of autonomy.

3.3. Granger Causality and TE

As discussed in section 2.5, GC requires stationarity, i.e., the mean and variance of each time series do not change over time), and that it can be adequately described by a linear model. We did the Dickey-Fuller test to check for stationary, and it affirmed that all signals (motor activation, velocities, accelerations) are stationary, but not all are linear, especially velocities and accelerations (as both have x and y components). Although GC itself is a linear measure, other non linear versions of causality have also been introduced (Marinazzo et al., 2008). Like standard (linear) GC, non-linear Granger measures are also potentially prone to overfitting and finding false positives. The kernel-based non-linear GC measure in a study by Marinazzo et al. (2008) argues that it solves these two issues and may potentially be a useful measure for this analysis. Similarly, quantum probability methods have been used for cognitive modeling, making extensive use of techniques similar to the

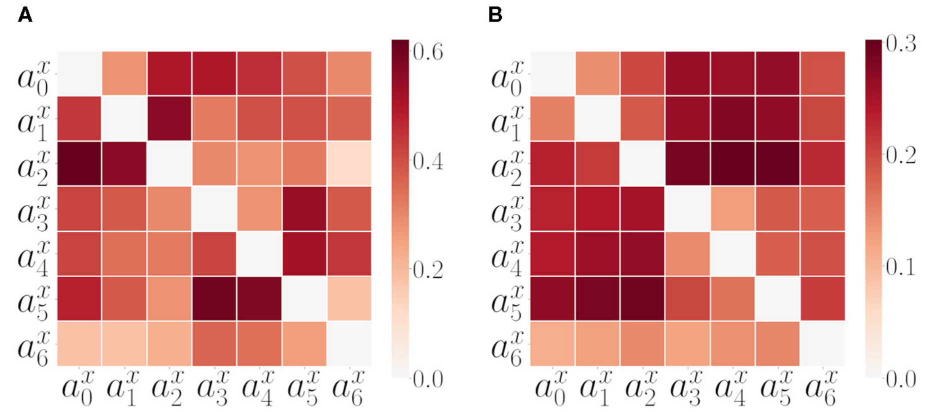


FIGURE 7 | Alternative measures: GC and TE. **(A)** Matrix of pairwise GC measure values among all SIPs for all SM interaction data. The matrix is computed considering the proportion of GC tests passed (refer to text for further details, section 3.3). **(B)** Matrix of pairwise TE measure values among all SIPs for all SM data.

kernel-based GC method (see Pothos and Busemeyer, 2013 for a general overview). However, besides the linear/non linear issue, GC also has another problem, that it is inadequate for situations where there are instantaneous effects, as is the case for our experiment here. This is further elaborated on pages 207–208 by Peters et al. (2017): "Knowing that a system contains instantaneous effects may suggest modifying GC by regressing not only on the past, but adding the current value, possibly leading to wrong bidirectional casual influences." This is in line with our observations.

With respect to TE (Vicente et al., 2011), it is advantageous when the model assumption of GC does not hold, for example, analysis of non linear signals. Thus, it makes sense as a logical next step from GC; however, TE still requires a lot of data and is not suitable for real-time inference. There are suggestions for significantly faster measures of TE based on permutation entropy (Bandt and Pompe, 2002). Although this method may improve over TE, it is not immediately clear whether it can outperform the AT measure, as it requires performing permutation analysis and comparison which is computationally expensive.

Two different sets of experiments were carried out to compare AT, Granger Causality (GC) and TE. In the following paragraphs, the results of the comparison experiments using the acceleration values of the SIPs are summarized[2]. In the first set of experiments, both GC and TE values were calculated on the generated signals explained in section 3.2. In the second set of experiments, all three measures were treated as binary classifiers of directional causality, and their performance tested on the experimental setup described in section 2.2 with increasingly more complex arm policy combinations.

Figure 4 summarizes the results of the first set of experiments, comparing the results of each measure for each signal set. For the first four signals, GC returns a p-value below 0, indicating x is Granger causing y, while for the last two signals GC did not find an effect. TE measure did not find an effect for the first and third signals, where the signals are not lagged, but does find a small entropy transfer for the lagged second and fourth signals regardless of the signal signs. TE finds strong effects in the last two signals, where the signals are random noises. The last version of the proposed AT measure only finds an effect on signals with opposite signs, Signal 1, 2, and 6, effect size decreasing with the lag amount.

Before executing the second set of experiments, we tested GC and TE on the experimental setup explained in Section 2.2 using a random policy for both arms. In **Figure 7** we plot the GC and Transfer Entropy matrices of all pairwise acceleration signals. Both are computed from the same SM data as in **Figure 6** extracted from 200 repetitions of a combined random behavior (as discussed in previous section 3.2). The GC matrix that we call, in this study, GC^{mat} (**Figure 7A**) is computed considering the proportion of passed GC tests (whenever $\mathrm{GC}(a_i^x, a_j^x)$ returned a $p < 0.05$). Part of the pattern present in the last column of the AT matrix is maintained in the GC matrix. Also the within arm dependencies are well captured, but we are not interested on those as they are already captured by the correlation analysis. In addition, we can also look at the directional dependencies by comparing $GC^{mat}(a_i^x, a_j^x)$ to $GC^{mat}(a_j^x, a_i^x)$, but here again, we got inconsistent results. TE matrix (**Figure 7B**) consistently gives a higher transfer in the direction from the arms to the object, with strong effects for the inter-arm dependencies.

Only considering the case with two random policy arms did not give enough information to make a fair comparison between the measures, leading to the second set of experiments. To assess the performance and limitations of each measure, seven scenarios with increasingly complex policy combinations were prepared:

- Scenario 1: Constant push and static
- Scenario 2: Staggered push and static
- Scenario 3: Random push and static
- Scenario 4: Random and static
- Scenario 5: Random and constant push
- Scenario 6: Random push and constant push
- Scenario 7: Random and random

---

[2] For GC we used the "statsmodels" Python module and for Transfer Entropy the Python interface to the Java Information Dynamics Toolkit.

The constant push policy consists of the arm pushing toward the ball with a constant motor activation, while the staggered push policy consists of the arm either moving with a constant motor activation or no motor activation with 0.5 probability each. The random push policy means the arm moves with variable motor activations sampled from a uniform distribution between 0 and 1. The random policy is the arm moving with motor activations sampled from a uniform distribution between –1 and 1. The static policy means that the motors are not activated at all and the arm does not move by itself. These policy combinations, although not exhaustive, represent a variety of different scenarios an agent can face in this setting. Haptic signals are used as the ground truth to determine which arms push the ball in each trial. Each scenario was executed for 100 trials, with the policies alternating between the arms each trial. The reports of each measure between the six SIPs of the arms and the SIP of the object were recorded and compared to the ground truth. As the accuracy score does not take into account false positives, and the previous set of experiments on the synthetic signals show that all the measures tend to report false positives in random signals, the F1 scores are used for the assessment. The F1 scores of the measures for each scenario are presented in **Table 1**.

The F1 scores in **Table 1** show that GC fails to correctly identify the direction of the causal relationships, especially with complex policies, reliably. This result is in line with expectations, as the setting pushes the limitations of the GC measure, explained at the beginning of the current section. TE and the AT measure successfully capture the directional causal relationships between the SIPs of the arms and the ball. AT keeping a stable performance across the first four scenarios where one arm is static, while the TE has decreased performance as the policy of the pushing arm gets more complex. Both measures have decreased performance in Scenarios 5 and 7, the only two scenarios where the arm that is not pushing the ball can have positive and negative acceleration and thus lead the measures to false positives as shown in the last signal in **Figure 4**.

In conclusion, correlational analysis of the SIPs can be used to characterize the self vs. the other and capture proximo-distal arm-joint dependencies and motor-sensory influences but is not sufficient to infer autonomy. TE, and the novel AT measure, but not GC, can reliably capture directional causalities between SIPs in this setting.

TABLE 1 | F1 scores of Acceleration Transfer (AT), Granger Causality (GC), and Transfer Entropy (TE) measures on seven mixed policy scenarios.

	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6	Scenario 7
AT F1 Score	0.91	0.91	0.9	0.89	0.78	0.91	0.77
GC F1 Score	0.67	0.41	0.16	0.27	0.56	0.21	0.17
TE F1 Score	0.91	0.84	0.48	0.8	0.68	0.82	0.69

4. DISCUSSION

We discuss, in this study, the minimal requirements for self/other distinction that lead to simpler methods in comparison with the approaches that require a multitude of parameters (such as deep recurrent neural networks Rabinowitz et al., 2018) or rely on the predictive coding hypothesis which requires a forward model to operate (Fairhurst et al., 2019). The typical approach to address the problem of self vs. other distinctions is based on the mismatch of the predicted perceptual state (given the precedent executed action) and the actual perceived state. A smaller mismatch of the (feedforward) prediction with the actually perceived state (feedback) would imply an increased degree of agency as stated by the *Comparator Model* (Wolpert and Flanagan, 2001; Farrer and Frith, 2002; Fairhurst et al., 2019).

We argue for the assumption that the ideal time for the distinction of self and other in early development to occur is prior to the acquisition of forward models, to be able to focus the learning and reduce the input dimensionality of the forward models to be learned. Following this assumption, we address the self and other distinction with the simplest method possible, a simple correlational analysis of visual movement features and in the absence of a forward model (Wolpert and Flanagan, 2001), or one could say that it is a forward model of purely visual features (as the introduced sensory to sensory predictions Maffei et al., 2017) and in a preliminary stage of learning to control. This preliminary learning phase is important as it can guide and reduce the dimensionality of the inputs to the more complex forward models (including motor signals) that subsequently need to be learned.

To clarify this simple mechanism, let us consider the following thought experiment. When playing a video game for the first time, we face a completely novel situation in which a forward model is not available. However, we can quickly discern which entities we are in control of by pressing buttons randomly, and we can direct attention to the immediate surroundings of those entities and reduce the state space dimensionality of the learning. In addition, imagine after having learned to control a character, we invert the SM mapping (pressing right makes the agent go left and vice-versa). The forward model that was acquired now makes wrong predictions (and needs to be relearned) but the feeling of self and other remained intact and has not been disrupted. Our approach, in this study, builds on Sánchez-Fibla et al. (2017b) and addresses exactly this preliminary stage.

We test this *via* a statistical and information theoretical analysis of the SM data stream (as performed in Hoffmann, 2014). The results show that the self/other distinction can be addressed solely by a correlational analysis of motor signals and their sensory effects (channeled through the motion, velocities and accelerations of the previously mentioned SIPs) prior to the construction of a forward model. For assessing autonomy, we need to go beyond correlation and perform a causality analysis to be able to extract directional dependencies. A rectified subtraction of feature accelerations, denoted in this study as AT, is shown to be sufficient to extract directional dependencies and as a cheaper alternative to more computationally expensive measures such as TE.

Autonomous motion alone is not the only information that infants use to assess agent/object autonomy as they check for autonomous control over the actions of entities' (Baillargeon et al., 2015). Considering the latter, there is a need to build forward models and the aforementioned *Comparator Model* applied to the assessment of control of other entities could be an explanation behind figuring out autonomy in its final phase. But before that, building a hierarchical structure of visual motion cues based on velocity correlations and AT may be a prerequisite step, as we have shown. From a neurophysiologycal perspective, there is evidence that there are neurons (in the brain MT area of the Macaque monkey) that are tuned to acceleration changes (Schlack et al., 2007).

5. CONCLUSIONS

From early development, self and other distinctions are fundamental to the focus the learning (reduce state space dimensionality) of forward model acquisition. From a computational perspective, discriminating between self and other features from visual feedback is often addressed through models (Brody et al., 2017; Sánchez-Fibla et al., 2017b; Thomas et al., 2017; Pertsch et al., 2018; Rybkin et al., 2018), which either require a multitude of parameters (deep learning approaches like Rabinowitz et al., 2018) or rely on the predictive coding hypothesis, requiring a forward model to be able to check the matching between current and predicted states (Fairhurst et al., 2019). In this study, we have approached this problem from a principled perspective, identifying minimum requirements to solve the problem of deciphering which features of the visual scene correspond to the self and which of them correspond to other entities in the scene, *via* a correlation analysis of velocity signals, that we have found to be sufficient. Thus, self/other distinctions could be identified with minimal and simpler methods, prior to the acquisition of the forward models and could guide and reduce the dimensionality of their inputs.

We do not work with images directly. Elements of the visual scene are interpreted *via* the SIPs, visual features (that we introduced in Sánchez-Fibla et al., 2017a, reminiscent of BM pivots that can be computed by computer vision methods like SIFT features). SIPs can be characterized as belonging to oneself thanks to the high motor to sensory correlations. Furthermore, from the full correlation matrix, proximo-distal joints can be characterized and also other entities with similar correlation patterns with their proximo-distal structures.

We go beyond the distinction of self and other by defining autonomy as the ability to move by oneself (animacy), and we discuss how we can detect it from visual SM interactions, a problem that has not been addressed yet, to our knowledge, from a computational perspective. For detecting autonomy, correlation comes with its limitations: it only captures linear relationships and does not allow us to infer directed/causal dependencies. To surpass this limitation, we looked at measures, such as GC, Copula-Based (Junker et al., 2019) and TE (Vicente et al., 2011), but we concluded that they were unable to capture directional dependencies on a trial-by-trial basis. Instead, they work better on distribution of sampled SM data. We grounded dependency assessment on a simpler principle of energy transfer between entities (energy in terms of acceleration). For this purpose, we developed a novel AT measure, that is not temporally dependent, and computed the estimated instantaneous transfer of acceleration (note, as we have discussed, that instantaneous effects pose problems for standard measures, such as GC) between two moving entities (in our case SIPs). The proposed AT measure works under the principle that an inanimate entity is always the sink in the directed SM graph of transfers and produces better results than the standard causality algorithms, but further research might compare other causal inference approaches (as discussed in section 3.3). AT would be a very natural way to interpret interactions between visual pivots as acceleration is proportional to force *via* Newton's 2nd law. In addition, neurophysiology findings back this hypothesis with proof of the existence of neurons with responses that are tuned to acceleration and deceleration (Schlack et al., 2007).

We have identified the minimal principles that we hypothesize are at play when making sense of embodied SM visual experiences, and we make a concrete proposal of what is the minimal level at which (causal) directional reasoning is needed to understand visual motion pivot interactions [as discussed in the challenges exposed in Pezzulo et al. (2011)]. Beyond the self/other distinction, which we have shown can be assessed by a correlational analysis of velocities (without the need for directional reasoning), we show that for the assessment of autonomy, directional inferences need to be utilized. We also hypothesize that these findings, grounded on developmental psychology, could also be transferred to developmental robotics (Cangelosi and Schlesinger, 2015). We also argue that transfer learning, that is the generalization of acquired knowledge from one task to another cannot be achieved without this fundamental step, by annotating the SM memory of the agent with "who did what and when." As proof of concept, we refer to the results obtained in the article by Demirel and Sánchez-Fibla (2019), where a reinforcement learning agent speeds up its learning by having access to the features it controls from its perceptual state. In this sense, approaches based on cognitive architectures will require this information in SM memory. Up until now, this has been lacking in the proposed architectures (refer to the mentioned DAC framework Duff et al., 2011 and the SM graph structures introduced in Toussaint, 2006).

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://github.com/santmarti/PythonRobot2DSim.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work was supported by INSOCO-DPI2016-80116-P, socSMC-641321—H2020-FETPROACT-2014, and FI-AGAUR 2020 FI_B 00693. This research was partially funded by the French National Research Agency (https://anr.fr), project ECOCURL, Grant ANR-20-CE23-0006, as well as the Inria Exploratory action ORIGINS (https://www.inria.fr/en/origins).

REFERENCES

Arsiwalla, X. D., Herreros, I., Moulin-Frier, C., Sanchez, M., and Verschure, P. (2016). "Is consciousness a control process?," in *Artificial Intelligence Research and Development* (IOS Press), 233–238.

Arsiwalla, X. D., Herreros, I., Moulin-Frier, C., and Verschure, P. (2017a). "Consciousness as an evolutionary game-theoretic strategy;" in *Conference on Biomimetic and Biohybrid Systems*, (Stanford, CA: Springer), 509–514.

Arsiwalla, X. D., Sole, R., Moulin-Frier, C., Herreros, I., Sanchez-Fibla, M., and Verschure, P. (2017b). The morphospace of consciousness. *arXiv Preprint arXiv*:1705.11190. doi: 10.31234/osf.io/emhnp

Baillargeon, R., Scott, R. M., He, Z., Sloane, S., Setoh, P., Jin, K.-,s., et al. (2015). Psychological and sociomoral reasoning in infancy. *Am. Psychol. Assoc.* doi: 10.1037/14341-003

Baker, C., Saxe, R., and Tenenbaum, J. (2011). "Bayesian theory of mind: Modeling joint belief-desire attribution," in *Proceedings of the Annual Meeting of the Cognitive Science Society, Vol* (Boston, MA). 33.

Bandt, C., and Pompe, B. (2002). Permutation entropy: a natural complexity measure for time series. *Phys. Rev. Lett.* 88:174102. doi: 10.1103/PhysRevLett.88.174102

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

Brody, J., Levin, S., and Shapiro, S. (2017). "Learning knowledge of agency," in *2017 IEEE International Conference on Autonomic Computing (ICAC)* (Columbus, OH: IEEE), 263–270.

Cangelosi, A., and Schlesinger, M. (2015). *Developmental Robotics: From Babies to Robots*. Cambridge, MA: MIT Press.

Demirel, B., and Sánchez-Fibla, M. (2019). "Speeding up reinforcement learning by combining attention and agency features," in *CCIA. Frontiers in Artificial Intelligence and Applications* (Mallorca), 84–94.

Duff, A., Fibla, M. S., and Verschure, P. F. (2011). A biologically based model for the integration of sensory-motor contingencies in rules and plans: a prefrontal cortex based extension of the distributed adaptive control architecture. *Brain Res Bull.* 85, 289–304. doi: 10.1016/j.brainresbull.2010.11.008

Fairhurst, M. T., Janata, P., and Keller, P. E. (2019). Distinguishing "self" from "other" in a dynamic synchronization task with an adaptive virtual partner. *BioRxiv* 625061. doi: 10.1101/625061

Farrer, C., and Frith, C. D. (2002). Experiencing oneself vs another person as being the cause of an action: the neural correlates of the experience of agency. *Neuroimage* 15, 596–603. doi: 10.1006/nimg.2001.1009

Freire, I. T., Arsiwalla, X. D., Puigbò, J., and Verschure, P. F. M. J. (2019). Modeling theory of mind in multi-agent games using adaptive feedback control. *CoRR, abs*/1905.13225.

Freire, I. T., Moulin-Frier, C., Sanchez-Fibla, M., Arsiwalla, X. D., and Verschure, P. (2018). Modeling the formation of social conventions in multi-agent populations. *arXiv preprint arXiv*:1802.06108.

Gelman, S. A. (2003). *The Essential Child: Origins of Essentialism in Everyday Thought*. Oxford: Oxford Series in Cognitive Dev.

Hoffmann, M. (2014). "Minimally cognitive robotics: body schema, forward models, and sensorimotor contingencies in a quadruped machine," in *Contemporary Sensorimotor Theory* (Cham: Springer), 209–233.

Junker, R. R., Griessenberger, F., and Trutschnig, W. (2019). A copula-based measure for quantifying asymmetry in dependence and associations. *arXiv preprint arXiv*:1902.00203.

Lowe, D. G. (1999). "Object recognition from local scale-invariant features," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference, Vol. 2* (Kerkyra: IEEE), 1150–1157.

Luo, Y., Kaufman, L., and Baillargeon, R. (2009). Young infants' reasoning about physical events involving inert and self-propelled objects. *Cogn. Psychol.* 58, 441–486. doi: 10.1016/j.cogpsych.2008.11.001

Maffei, G., Herreros, I., Sanchez-Fibla, M., J., Friston, K., and, F. M. J., et al. (2017). The perceptual shaping of anticipatory actions. *Proc. R. Soc. B Biol. Sci.* 284:20171780. doi: 10.1098/rspb.2017.1780

Marinazzo, D., Pellicoro, M., and Stramaglia, S. (2008). Kernel method for nonlinear granger causality. *Phys. Rev. Lett.* 100:144103. doi: 10.1103/PhysRevLett.100.144103

Marsella, S. C., Pynadath, D. V., and Read, S. J. (2004). "Psychsim: Agent-based modeling of social interactions and influence," in *Proceedings of the International Conference on Cognitive Modeling, Vol* (Pittsburgh, PA). 36, 243–248.

Mascalzoni, E., Regolin, L., and Vallortigara, G. (2010). Innate sensitivity for self-propelled causal agency in newly hatched chicks. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4483–4485. doi: 10.1073/pnas.0908792107

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., et al. (2015). Human-level control through deep reinforcement learning. *Nature* 518, 529. doi: 10.1038/nature14236

Nematzadeh, A., Burns, K., Grant, E., Gopnik, A., and Griffiths, T. L. (2018). Evaluating theory of mind in question answering. *arXiv preprint arXiv*:1808.09352. doi: 10.18653/v1/D18-1261

Opfer, J. E., and Gelman, S. A. (2011). "Development of the animate-inanimate distinction," in *The Wiley-Blackwell Handbook of Childhood Cognitive Development* (Hobokon, NJ), Vol. 2, 213–238.

Pavlova, M. A. (2011). Biological motion processing as a hallmark of social cognition. *Cereb. Cortex* 22, 981–995. doi: 10.1093/cercor/bhr156

Pertsch, K., Rybkin, O., Derpanis, K., and Jaegle, A. (2018). "Unsupervised discovery of an agent's action space via variational future prediction," in *NeurIPS Workshop Infer2Control on Probabilistic Reinforcement Learning and Structured Control* (Montréal, QC).

Peters, J., Janzing, D., and Schölkopf, B. (2017). *Elements of Causal Inference*. Cambridge, MA: The MIT Press.

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., Spivey, M., and McRae, K. (2011). The mechanics of embodiment: a dialog on embodiment and computational modeling. *Front. Psychol.* 2:5. doi: 10.3389/fpsyg.2011.00005

Pothos, E. M., and Busemeyer, J. R. (2013). Can quantum probability provide a new direction for cognitive modeling? *Behav. Brain Sci.* 36, 255–274. doi: 10.1017/S0140525X12001525

Rabinowitz, N. C., Perbet, F., Song, H. F., Zhang, C., Eslami, S., and Botvinick, M. (2018). Machine theory of mind. *arXiv preprint arXiv*:1802.07740.

Razak, F. A., and Jensen, H. J. (2014). Quantifying 'causality' in complex systems: understanding transfer entropy. *PLoS ONE.* 9:e99462. doi: 10.1371/journal.pone.0099462

Rybkin, O., Pertsch, K., Derpanis, K. G., Daniilidis, K., and Jaegle, A. (2018). Learning what you can do before doing anything. *arXiv preprint arXiv*:1806.09655.

Sánchez-Fibla, M., Duff, A., and Verschure, P. F. (2011). "The acquisition of intentionally indexed and object centered affordance gradients: a biomimetic controller and mobile robotics benchmark," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference* (San Francisco, CA: IEEE), 1115–1121.

Sánchez-Fibla, M., Forestier, S., Moulin-Frier, C., Puigbò, J.-Y., and Verschure, P. F. (2019). From motor to visually guided bimanual affordance learning. *Adaptive Behavior*, 28, 63–78. doi: 10.1177/1059712319855836

Sánchez-Fibla, M., Maffei, G., and Verschure, P. F. M. J. (2017a). "Adaptively learning levels of coordination from one's, other's and task related errors through a cerebellar circuit: a dual cart-pole setup," in *Artificial Neural Networks and Machine Learning-ICANN 2017*, eds A. Lintas, S. Rovetta, P. F. Verschure and A. E. Villa (Cham: Springer International Publishing), 309–316.

Sánchez-Fibla, M., Moulin-Frier, C., Arsiwalla, X. D., and Verschure, P. F. (2017b). "Social sensorimotor contingencies: towards theory of mind in synthetic agents," in *CCIA. Frontiers in Artificial Intelligence and Applications* (Tarragona), 251–256.

Schlack, A., Krekelberg, B., and Albright, T. D. (2007). Recent history of stimulus speeds affects the speed tuning of neurons in area mt. *J. Neurosci.* 27, 11009–11018. doi: 10.1523/JNEUROSCI.3165-07.2007

Tager-Flusberg, H. (2007). Evaluating the theory-of-mind hypothesis of autism. *Curr. Dir. Psychol. Sci.* 16, 311–315. doi: 10.1111/j.1467-8721.2007.00527.x

Thomas, V., Pondard, J., Bengio, E., Sarfati, M., Beaudoin, P., Meurs, M.-J., et al. (2017). Independently controllable features. *arXiv preprint arXiv*:1708.01289.

Toussaint, M. (2006). A sensorimotor map: Modulating lateral interactions for anticipation and planning. *Neural Comput*. 18, 1132–1155. doi: 10.1162/neco.2006.18.5.1132

Verschure, P. (2016). Synthetic consciousness: the distributed adaptive control perspective. *Philos. Trans. R. Soc. B Biol. Sci.* 371:20150448. doi: 10.1098/rstb.2015.0448

Vicente, R., Wibral, M., Lindner, M., and Pipa, G. (2011). Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *J. Comput. Neurosci.* 30, 45–67. doi: 10.1007/s10827-010-0262-3

Wolpert, D. M., and Flanagan, J. R. (2001). Motor prediction. *Curr. Biol.* 11, R729–R732. doi: 10.1016/S0960-9822(01)00432-8

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Human Neuroscience

HYPOTHESIS AND THEORY
published: 15 September 2021
doi: 10.3389/fnhum.2021.624610

Socializing Sensorimotor Contingencies

*Annika Lübbert[1], Florian Göschl[1], Hanna Krause[2], Till R. Schneider[1], Alexander Maye[1] and Andreas K. Engel[1]**

[1] Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany,
[2] Department of Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

The aim of this review is to highlight the idea of grounding social cognition in sensorimotor interactions shared across agents. We discuss an action-oriented account that emerges from a broader interpretation of the concept of sensorimotor contingencies. We suggest that dynamic informational and sensorimotor coupling across agents can mediate the deployment of action-effect contingencies in social contexts. We propose this concept of *socializing sensorimotor contingencies* (socSMCs) as a shared framework of analysis for processes within and across brains and bodies, and their physical and social environments. In doing so, we integrate insights from different fields, including neuroscience, psychology, and research on human–robot interaction. We review studies on dynamic embodied interaction and highlight empirical findings that suggest an important role of sensorimotor and informational entrainment in social contexts. Furthermore, we discuss links to closely related concepts, such as enactivism, models of coordination dynamics and others, and clarify differences to approaches that focus on mentalizing and high-level cognitive representations. Moreover, we consider conceptual implications of rethinking cognition as social sensorimotor coupling. The insight that social cognitive phenomena like joint attention, mutual trust or empathy rely heavily on the informational and sensorimotor coupling between agents may provide novel remedies for people with disturbed social cognition and for situations of disturbed social interaction. Furthermore, our proposal has potential applications in the field of human–robot interaction where socSMCs principles might lead to more natural and intuitive interfaces for human users.

Keywords: sensorimotor contingencies, coupling, prediction, human–robot interaction, coordination dynamics, joint action, autism spectrum disorder

Edited by:
Leonhard Schilbach,
Ludwig Maximilian University
of Munich, Germany

Reviewed by:
Stephanie D. Preston,
University of Michigan, United States
Giovanni Di Pino,
Campus Bio-Medico University, Italy

***Correspondence:**
Andreas K. Engel
ak.engel@uke.de

Specialty section:
This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

Received: *31 October 2020*
Accepted: *24 August 2021*
Published: *15 September 2021*

Citation:
Lübbert A, Göschl F, Krause H,
Schneider TR, Maye A and Engel AK
(2021) Socializing Sensorimotor
Contingencies.
Front. Hum. Neurosci. 15:624610.
doi: 10.3389/fnhum.2021.624610

INTRODUCTION: GROUNDING COGNITION IN ACTION

In recent years, a ‘pragmatic turn’ has been emerging in the cognitive sciences, i.e., a conceptual move away from the classical representation-centered framework toward a paradigm that emphasizes the close relation between cognition and action (for review, see Engel et al., 2013b, 2016). Although such an action-oriented paradigm has been supported by many proponents over the years (e.g., Varela et al., 1991; Clark, 1997; Noë, 2004), it has only recently begun to show conspicuous impact in the cognitive sciences (see Engel, 2010; Menary, 2010; Engel et al., 2013b; Durt et al., 2017). The basic notion is that cognition should not be conceived as the capacity of compiling world-models, which then provide a detached database for independent thinking, planning, and problem solving (Schilbach et al., 2013). Rather, it is emphasized that cognitive processes are so closely intertwined with a body in action that cognition is best understood as enactive, as a form of situated practice rather than disembodied mentalizing (Varela et al., 1991; Noë, 2004; Engel, 2010). Cognition, in this view, is grounded in a pre-rational being-in-the-world based on sensorimotor skills for real-life situations, and core aspects of cognition, such as sensing, perceiving or understanding, become inseparable from doing (Varela et al., 1991; Clark, 1997; O'Regan and Noë, 2001; Noë, 2004). This agrees with phenomenological claims about intricate links between our different senses and the body's role in thinking (Merleau-Ponty, 1962, 1963), modern anthropological studies of the process of knowledge-making (Myers and Dumit, 2011; Myers, 2015) and recent calls to look beyond analytic ways of knowing (De Jaegher, 2019). Inspired by pragmatist and phenomenological traditions, numerous recent authors have explored the implications of defining cognition as embodied action (Varela et al., 1991; Clark, 1997; Noë, 2004; Pfeifer and Bongard, 2006; Engel, 2010; Menary, 2010; Sheets-Johnstone, 2011; Engel et al., 2013b).

Immediate precursor to the concept proposed in this article, the 'sensorimotor contingency theory' (SMCT) by O'Regan and Noë (2001) centers on the notion that perception and cognition can only be understood by considering their inherent action-relatedness. In this framework, sensorimotor contingencies (SMCs) are defined as acquired law-like relations between movements and associated changes in sensory inputs that are continuously probed and refined as we orient in the world (O'Regan and Noë, 2001). The formation of SMCs shows to be highly relevant in cognition (O'Regan and Noë, 2001; Engel et al., 2013b; Maye and Engel, 2013). SMCs are acquired through the agent's actions, and are deemed constitutive for perceptual processes. For instance, according to the SMCT seeing cannot be understood as computation on internal visual representations. Rather, seeing corresponds to engagement in visual exploratory activity, and consists in sets of skills that are mediated by knowledge in the form of SMCs. This active nature of perception has been emphasized by other approaches as well. However, the concept of SMCT is more radical: it considers action a necessary prerequisite for perception, not just as an output capacity that supports, or interacts with, perceptual processing. Of note, this account does not postulate a unidirectional impact of motor systems on perception but, rather, is compatible with the notion of dynamic sensorimotor interactions in reentrant processing loops (Engel, 2010). There is increasing evidence from work in neuroscience, psychology and robotics supporting the SMCT perspective (e.g., Frith et al., 2000; Maravita and Iriki, 2004; Gallese and Lakoff, 2005; Schubotz, 2007). For instance, neuronal response properties in sensory brain regions strongly depend on action context (Gallant et al., 1998), perceptual scene segmentation is facilitated by the active use of the objects (Bergström et al., 2011), and processes like attention and decision-making have been shown to be strongly related to activity of motor regions (Moore et al., 2003; Donner et al., 2009). Thus, SMCs have been proposed as a framework to define object concepts and action plans, suggesting that the mastery of sensorimotor contingencies facilitates goal-oriented behavior (Maye and Engel, 2011, 2012; Engel et al., 2013b; Högman et al., 2013). This implies that SMCs can be relevant over variable time scales beyond the correlation between movements and the immediate changes in sensory inputs, which are the focus of the original SMCT (O'Regan and Noë, 2001).

In keeping with this pragmatic turn, the concept discussed here suggests an action-oriented framework for social cognition in biological and artificial agents. Our proposal is to ground even complex modes of social interaction in the continuous dynamic coupling between agents and their environments. Successful social interaction, thus, does not come about exclusively through the theories that a detached observer holds about the intentions, beliefs and personalities of other agents (Carruthers and Smith, 1996) but – as we will argue – to a substantial extent via the formation and management of shared rhythms and patterns at the level of embodied sensorimotor dynamics. As will be discussed in greater detail below, our proposal is related to and inspired by other action-oriented concepts of social cognition that have emphasized the relevance of coordination dynamics (Tognoli and Kelso, 2014), of socially salient movement patterns (Lindblom and Ziemke, 2006), motor mimicry (Wang and Hamilton, 2012) and joint embodied action (Sebanz et al., 2006). Notably, earlier proponents of an enactive view of social cognition have suggested that even complex types of social interactions may be grounded in basic sensorimotor patterns that enable the dynamic coupling of agents (De Jaegher et al., 2010, 2017). Supporting this view, evidence is available that interactive sensorimotor dynamics provide substantial clues to social understanding (Di Paolo and De Jaegher, 2012), give rise to high-level processes such as shared intentionality (Sebanz et al., 2006) and empathy (De Waal and Preston, 2017), and are highly relevant for interpersonal affiliation, trust and prosocial behavior (Keller et al., 2014).

In the concept proposed here, the notion of SMCs is substantially broadened beyond its original scope (O'Regan and Noë, 2001) to include the learning and deployment of action-effect predictions on longer time-scales and more complex levels of processing. Previously, we have suggested that SMCs can be deployed, for instance, to acquire object concepts and to achieve prediction and action planning (Maye and Engel, 2011, 2013). Here, we propose that the relevance of SMCs is not limited to cognitive processing of the individual, but extends into the effective interactions between agents in social context. Since in our view these socially shared contingencies are constitutive for social cognition, the influence of others cannot be discarded when seeking to explain individual cognition or behavior: individual and collective processes become irreducibly linked. In the following, we use 'socializing sensorimotor contingencies (socSMCs)' as a shorthand for the proposal to ground the development and instantiation of social cognition in shared action-effect contingencies.

UNPACKING THE socSMCs CONCEPT

The socSMCs concept moves away from the classical notion that presumptive higher levels of cognition (e.g., self-recognition, perspective-taking, planning, complex reasoning) might differ fundamentally from presumed basic levels of sensorimotor processing (such as perception, multisensory integration, or motor coordination). This aligns well with the notion that both domains of cognition rely on common neural architectures and computational principles (Keller and Mrsic-Flogel, 2018), and evidence that brain regions embodying complex cognitive functions do not differ in principle from modules involved in more basic functions (Douglas and Martin, 2004). Where classical cognitivism might ask, 'How would we understand the world, other than by generating models about it?,' the socSMCs concept acknowledges the role of abstract reasoning, but puts equal emphasis on collective sense-making processes that arise only in relation to our physical and social environments. Thus, the socSMCs concept suggests in principle shared neural mechanisms for all our ways of engaging with our environment, and views structures and activities outside of our central nervous system as essential for our cognitive abilities (Clark and Chalmers, 1998).

A key assumption in the concept of socSMCs is that agents deploy learned action-effect contingencies in social contexts to anticipate outcomes of their own and others' actions (Brown and Brüne, 2012): I am the initiator of change in the (social) world, and change in the world can be directed at me. Such action-effect contingencies closely relate to the more basic framework of SMCs described above where, e.g., stable perception of the world comes about because we actively learn patterns of correlations between our actions (eye movements) and the ensuing effects (changes in the retinal inputs). We propose that agents' ability to anticipate and coordinate with others at linguistic and abstract levels may derive from their learning of motivated and embodied action in the world. In other words: how we orient in social contexts is very much an extension of how our body orients in the world. This includes social entrainment, defined by the sensorimotor or informational coupling between agents, and social engagement, i.e., the experience of connectedness or relatedness to other agents. The socSMCs concept predicts that both are grounded in the acquisition and deployment of action-effect contingencies. Further, we assume that both the experience of social engagement and our participation in social entrainment are situated within particular physiological, cultural and environmental contexts, within which they emerge and onto which they also feed back.

Another central assumption in the socSMCs concept is that social interaction can best be conceptualized in terms of dynamic coupling at different scales (Hasson et al., 2012; Engel et al., 2013a; Keller et al., 2014; Hasson and Frith, 2016; Kelso, 2019). We propose to differentiate three levels of complexity of social coupling, reflecting different stages across which interactions are established in a multi-agent system (**Figure 1**). We term these 'check SMCs,' 'sync SMCs,' and 'unite SMCs,' respectively, to denote that they may correspond to distinct stages, or modes, of social entrainment. These levels are best conceived as points on a continuum, with potential co-occurrence of modes of relating. Across these different levels of socially deployed SMCs, coupling is established over an increasing set of degrees of freedom of the interacting multi-agent system. At the first level, check SMCs involve unidirectional coupling, one agent predicting another agent's actions or the interaction between several other agents. Behaviorally, this may lead, e.g., to entrainment of one agent to a group of other agents. At the next level, sync SMCs enable bidirectional coupling, with both agents mutually sharing, attending to and predicting each other's sensorimotor actions. This reciprocity may then lead to genuine interactions and mutual entrainment of behavior, facilitating cooperation, joint attention, turn-taking, and shared action goals. At the third level, we suggest unite SMCs as a hypothetical coupling mode that may promote group-related, multidirectional coupling. Unite SMCs might be characterized by the emergence of interaction patterns that cannot fully be explained by the pairwise interactions among the group members, and attain a certain amount of autonomy over them (see also De Jaegher et al., 2017). For brain networks, there is evidence to suggest the occurrence of such higher-order coupling modes. Thus, it has been shown that cortical activity contains correlation patterns involving spikes from three or four neurons more often than predicted from pairwise correlations, and that such higher-order patterns relate to information encoding and behavior (Montani et al., 2009; Shimazaki et al., 2012). We hypothesize that similar higher-order dynamics might occur for social coupling modes. Such group dynamics may play a key role in group mental states, shared habits, and group affect. At this level, the emergent macroscopic pattern of multi-agent coupling may be stable enough to provide a new source of entrainment for individual agents, beyond the impact of pairwise interactions, as has been observed, e.g., in studies on collective dance improvisation (Himberg et al., 2018).

We suggest that these types of SMCs may take effect over different temporal and spatial ranges, depending on the setting and the mechanisms involved in the interaction. In this context, it may be useful to distinguish between 'proximal' and 'distal' interactions (**Figure 2**; cf. Pezzulo et al., 2019). While proximal interactions involve direct physical contact and sensorimotor coupling, distal interactions promote social entrainment by information flow between agents without direct physical coupling. Both proximal and distal social coupling abound in everyday life. Real-world scenarios involving proximal interactions with direct sensorimotor coupling include, for instance: greeting habits, like a handshake or a hug, where mutual dynamic entrainment is highly relevant for signaling the quality of a social relation; joint lifting or carrying of heavy objects that cannot be handled by one person alone, e.g., when moving a household; or dancing together as a couple, where sensorimotor coupling creates the synergy and togetherness enjoyed by the dancers. Examples for distal SMCs in social context include: social mimicry, i.e., an involuntary tendency

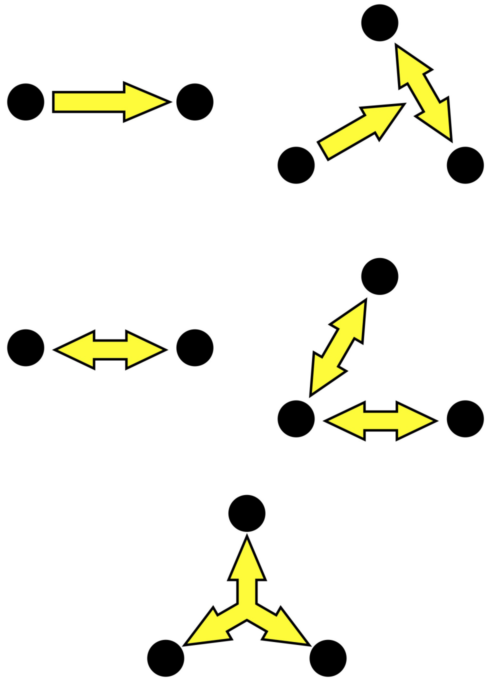


FIGURE 1 | Three hypothesized levels of SMCs in social interaction: **(Top)** Check SMCs may be mediated by unidirectional coupling between two agents (left) or from one person to other interacting agents (right). **(Middle)** Sync SMCs involve reciprocal coupling between two or more agents. **(Bottom)** Unite SMCs are conceived as emergent higher-order correlation patterns in the group dynamics.

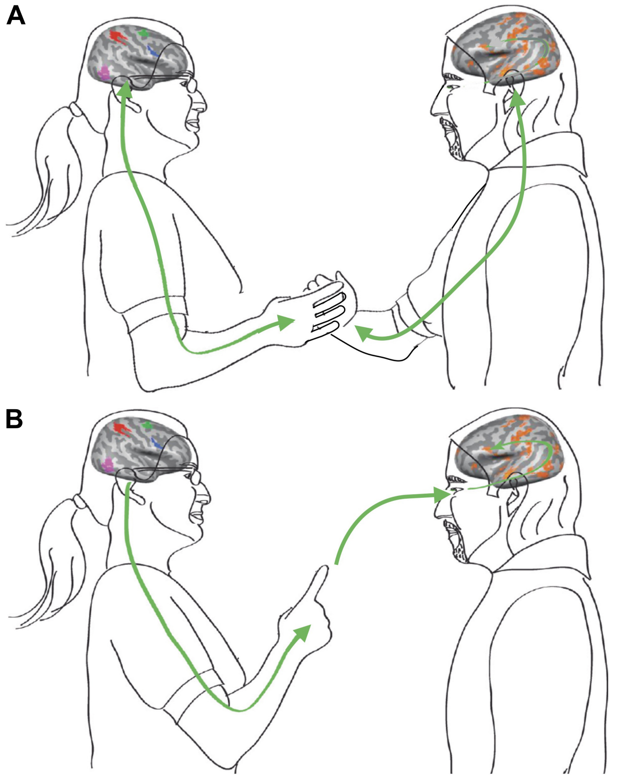


FIGURE 2 | Social interactions may involve proximal and distal types of SMCs. **(A)** Proximal sensorimotor coupling through direct physical contact, involving haptic sensing and kinesthesia. **(B)** Distal sensorimotor coupling based on distance senses including vision and audition to feed action-perception loops. Modified from Hasson et al. (2012).

to imitate or synchronize with postures and gestures of a conversation partner; team sports, ranging from synchronized swimming to coordinated group dynamics in volleyball or soccer; performance of musical ensembles engaged in joint improvisation, or the informational coupling between conductor and orchestra through embodied movement cues. Of note, distal interactions based solely on informational coupling can also take effect in fully virtual settings such as, e.g., in online gaming or in a video conference, provided that the agents can engage in meaningful action-effect contingencies.

The socSMCs concept treats individuals engaged in an interaction as one system. It therefore requires methods suited for the analysis of complex systems, since they may best capture the reciprocal adaptation that underlies coordination and communication (Fusaroli and Tylén, 2016; Gallotti et al., 2016). To this end, we suggest that measures used to quantify coupling within brains (for review, see Engel et al., 2013a) could prove equally useful to quantify the degree of coupling between individuals and their environment. Dynamic functional coupling is considered a key feature of neuronal activity, which exhibits rich spatiotemporal patterning and strongly modulates cognitive processing. Measures used to quantify functional coupling in the brain include coherence, power envelope correlation, information-theoretic measures or multivariate autoregressive models (see, e.g., Engel et al., 2013a; Hutchison et al., 2013; Bastos and Schoffelen, 2015). Much of this coupling is intrinsically generated, that is, not imposed by entrainment to an external stimulus or movement, but emerging from the connections within neuronal networks. There is clear evidence for two distinct types of coupling modes, which seem to be based on different coupling mechanisms (Siegel et al., 2012; Engel et al., 2013a). One type arises from phase coupling of band-limited oscillatory activity, whereas the other results from coupled aperiodic fluctuations of power envelopes. These two coupling modes (phase coupling vs. envelope coupling) differ in their dynamics, their spatial distribution, the time scales over which they operate and they likely support different functions (Engel et al., 2013a). Envelope coupling might reflect co-activation of regions on slower time scales and, thus, might facilitate the participation of brain areas in an upcoming task. Phase coupling, in contrast, represents coupling on faster time scales which presumably generates highly specific dynamic links within networks defined by envelope coupling. As part of the socSMCs concept, we propose that these intrinsic coupling modes are complemented by extrinsic coupling modes, i.e., coupling patterns that reflect the interaction of the brain with the body and its environment, including the social context (**Figure 2**; cf. Hasson et al., 2012; Hasson and Frith, 2016; Pezzulo et al., 2019). We propose that such extrinsic coupling modes may play a key role in enabling coordinated interaction of multiple brain systems with both body and environment, and that they may be particularly relevant for interaction with the social world. These extrinsic coupling modes should not only become evident at the level of behaviors or movement kinematics, but also give rise to inter-brain coupling in settings where neural signals can be concurrently recorded from two or more subjects (see section on 'Extrinsic neural coupling modes' below).

In summary, we suggest the notion of coupling with varying levels of complexity (check, sync and unite SMCs) and an integrated perspective of intrinsic and extrinsic coupling modes to be particularly helpful to understand social behavior. A key prediction is that changes of social entrainment, i.e., proximal or distal sensorimotor coupling, should be associated with changes in social engagement, which may be quantified by subjective ratings of the interaction quality or the degree of cooperation. Thus, we expect that a modulation of social coupling modes, in particular at the level of sync SMCs and unite SMCs, should lead to changes in presumed high-level social cognitive phenomena, such as mutual trust or empathy (Froese et al., 2014; Keller et al., 2014; Llobera et al., 2016). To achieve such a modulation, entrainment through shared perceptual and sensorimotor rhythms is likely to be an important mechanism. Conversely, fluctuations in social engagement might also lead to a differently organized dynamics of intrinsic and extrinsic coupling modes. Thus, for instance, the dynamics of sensorimotor coordination of two individuals should be influenced by social-cognitive factors such as shared intentionality or joint attention. Furthermore, the socSMCs concept emphasizes the continuity between low-level SMCs, which directly involve sensory and motor areas, as well as basal ganglia and cerebellum, and socially deployed action-effect contingencies. Thus, we hypothesize that there may be a strong overlap regarding the brain networks involved in both the former and the latter, as well as an interaction between the intrinsic and extrinsic coupling modes subserving the different types of SMCs. Moreover, with its focus on shared perceptual and sensorimotor rhythms as a core part of the architecture of social cognition, the socSMCs concept leads to the hypothesis that disturbances of these coupling modes may contribute to clinical deficits in social cognition, and that interventions at this level may provide an important tool to promote well-being at an interpersonal level.

RELATION TO OTHER CONCEPTS OF SOCIAL INTERACTION

According to the socSMCs concept, social interaction strongly depends on dynamic coupling between agents and their environment, hence a deeper understanding of this interaction dynamics promises to provide important insights into social cognition. Our view shares aspects with the interactionist concept of social cognition (Di Paolo and De Jaegher, 2012; Di Paolo et al., 2018; De Jaegher, 2019) which proposes an extension of the enactivist position to social and affective domains, emphasizing that sense-making occurs in a participatory way and that core aspects of cognition are inherently relational (De Jaegher and Di Paolo, 2007; De Jaegher et al., 2010; see also Durt et al., 2017). The proponents of this enactive view of social cognition emphasize the relevance of self-other contingencies for the coordination between agents in the interaction process (McGann and De Jaegher, 2009). However, a difference to the socSMCs concept is that a relation between social entrainment and intrinsic dynamics of the agents, in particular intrinsic neural coupling modes, is not considered. Furthermore, our concept agrees well with the joint action model by Knoblich and Sebanz (2008), which creates a close link between shared intentionality and joint action, based on the consideration of scenarios with different levels of complexity and flexibility of social interaction. However, the aspect of dynamic coupling is not considered in this model which, rather, focuses on the representation of perceived action in the agents (Sebanz et al., 2006; Knoblich and Sebanz, 2008).

Relations also exist to the concept of 'coordination dynamics,' which originated from earlier ideas on self-organizing pattern formation (Tognoli and Kelso, 2014; Tognoli et al., 2020). Coordination dynamics applies dynamical systems theory to biological networks, suggesting that a system is best described by looking at the coupling of its parts via mutual information exchange. An important distinction at the heart of this dynamical view is between (1) coupling of system components with similar dynamics, leading to formation of attractors or multistability; and (2) coupling of system parts with dissimilar dynamics, which prevents phase-locking and leads to metastability, i.e., integrative and segregative tendencies alternate in the interaction dynamics. Tognoli and Kelso (2014) have suggested that these two modes of coupling (multistable vs. metastable) might be useful to describe social coordination. Metastability is particularly interesting also because it represents a state of collective dynamics where new information can be created (Tognoli and Kelso, 2014). The application of this concept to the case of social interaction has been shown to provide very useful tools for the analysis of the interaction dynamics, such as coupled oscillator models (Tognoli et al., 2020). Yet, the focus of this approach has so far been on behavioral aspects of the coordination dynamics and not primarily on the explanation of social cognition and social perception.

Of note, the socSMCs concept differs from classical concepts in social neuroscience. A major focus of work on the neural foundations of social cognition has, in the past decades, been on the capacity of the brain to mirror the actions of others, thus enabling the simulation and representation of other agents' mental states (Gallese and Goldman, 1998). One of the highly interesting aspects of this approach is its strong emphasis on the role of motor and premotor systems in social cognition. Neuroimaging studies have identified brain areas and networks that are activated during tasks involving mentalizing, empathy or mirroring (Stanley and Adolphs, 2013). A relation between motor control and social cognition is also suggested by work on motor mimicry, an unconscious and spontaneous form of interpersonal coordination, which is likely mediated by the mirror neuron system (Wang and Hamilton, 2012). Along the same lines, De Waal and Preston have proposed a perception-action model of empathy, which postulates the emergence of empathy from basic sensorimotor processes and overlapping representations for performing and observing actions (De Waal and Preston, 2017). Several approaches have suggested a key role for predictive mechanisms in social cognition and also have explored their relevance for disturbed social processing (Blakemore and Decety, 2001; Brown and Brüne, 2012; Sinha et al., 2014). Sokolov et al. (2017) and Sokolov (2018) have highlighted the potential relevance of cerebellar circuits for signaling of prediction errors in social context. In contrast to the majority of the concepts that have been developed in social neuroscience so far, the socSMCs concept focuses on low-level sensorimotor interactions leading to social entrainment and engagement and, vice versa, the influence of social context on the development of basic sensorimotor relations. Pezzulo et al. (2019) emphasize the role of sensorimotor communication in social interaction scenarios of different complexity but without any link to the concept of sensorimotor contingencies. Hasson et al. (2012) and Hasson and Frith (2016) have proposed that social interactions involve the informational coupling of the perceptual system of one brain to the motor system of another which can lead to behavioral alignment, e.g., in verbal communication. However, these authors do not explicitly consider the link between such an extrinsic coupling to intrinsic coupling modes.

The socSMCs concept also differs from classical concepts in social cognition research, in particular, from theory of mind-based approaches. The concept of a theory of mind refers to the idea that a person is aware of the existence of their own subjective experience of the world, and the difference to that of another person. As such, research into this direction describes and promotes social interaction as mediated by theory-theory or simulation-theory (Carruthers and Smith, 1996; Gallese and Goldman, 1998), both of which invoke a meta-level of social cognition, and a distancing from the ongoing moment-to-moment interaction with other agents. In contrast, the socSMCs concept emphasizes the role of more basic and immediate processes of social sense-making, seeking to explain how abstract or higher level insights and decisions come about and are informed by bodily, dynamic and situational factors. This notion also aligns well with evidence from developmental research, suggesting that early in development, the social interaction modes emphasized in the socSMCs concept have primacy and are required to ground other, more explicit modes of social cognition (Campos et al., 2000; Di Paolo and De Jaegher, 2012). Rather than foregrounding models that we hold about others and our interactions with them, the socSMCs concept promotes a picture in which agents co-create shared effects in the world and, thus, understand sociality through the experience of enacting 'we-modes' (Varela et al., 1991; De Jaegher and Di Paolo, 2007; De Jaegher et al., 2017). It should be noted that both ways of knowing matter: cognitive model-based prediction and dynamic social coupling, both involve habitual as well as creative components, mutually influence one another and contribute to our flexible engagement with the world (see also Pezzulo et al., 2019). Nonetheless, given the frequent lack of intra- and interpersonal sensorimotor, and experientially lived aspects of cognition in representational approaches, the socSMCs concept is an invitation to keep abstract reasoning and embodied relating at par, acknowledging that the two ways of understanding rely on each other.

SOCIAL COORDINATION DYNAMICS

A major implication of the socSMCs concept is a shift in terms of what should be considered as core mechanisms of social cognition. How do we come to understand each other, work on a task together, or settle a dispute? According to the concept advocated here, for multiple agents to act together and understand one another, they must first and foremost find a way to coordinate their sensorimotor engagement with the world and with one another.

The importance of sensorimotor coordination for joint action is particularly evident in behaviors involving shared rhythms such as the applause of an audience which can occur in spontaneously emerging synchrony across many individuals. The dynamics of social coordination has been studied, for example, during rhythmic finger movements carried out by dyads of participants with and without visual feedback regarding their own and the other's movements (Oullier et al., 2008; **Figure 3**). In epochs with visual feedback, phase synchrony emerged spontaneously between the finger movements, although the participants had not received any particular instruction about how to relate to the partner's finger movements. Of note, the effect of social entrainment persisted after periods of phase synchronization when visual feedback was eliminated by closing the eyes (**Figure 3**). This study provides a typical example for what we have termed sync SMCs above (**Figure 1**). The authors conclude that general features of coordination dynamics, such as multistability and phase transitions, which are observed in a broad variety of self-organizing dynamical systems, are also highly relevant in social interaction. These conclusions are also supported by recent work on joint rushing, i.e., the unconscious increase in pace that can occur during synchronized rhythmic activities (Wolf et al., 2019).

Further prime examples for social entrainment are provided by the coordination dynamics among musicians during ensemble performance (reviewed by Keller et al., 2014). In contrast to more basic laboratory paradigms, entrainment in musical ensembles requires coordination of complex movement sequences with variable temporal patterning. It has been suggested that several cognitive and sensorimotor capacities are required for successful social coupling in such complex settings, including (i) temporal adaptation, supported by mechanisms such as phase correction and period correction; (ii) attention to both the results of own actions, actions of the partners and the joint ensemble output; and (iii) anticipation of action outcomes based on highly precise temporal prediction capabilities (van der Steen and Keller, 2013; Keller et al., 2014). These studies in musical ensembles provide evidence for an impact of sensorimotor coordination on social cohesion, cooperation and trust and, overall, they provide a highly relevant case where synchronous group entrainment can enhance social affiliation (D'Ausilio et al., 2015). Similar conclusions have been reached in the study of musical improvisation involving duets or larger ensembles

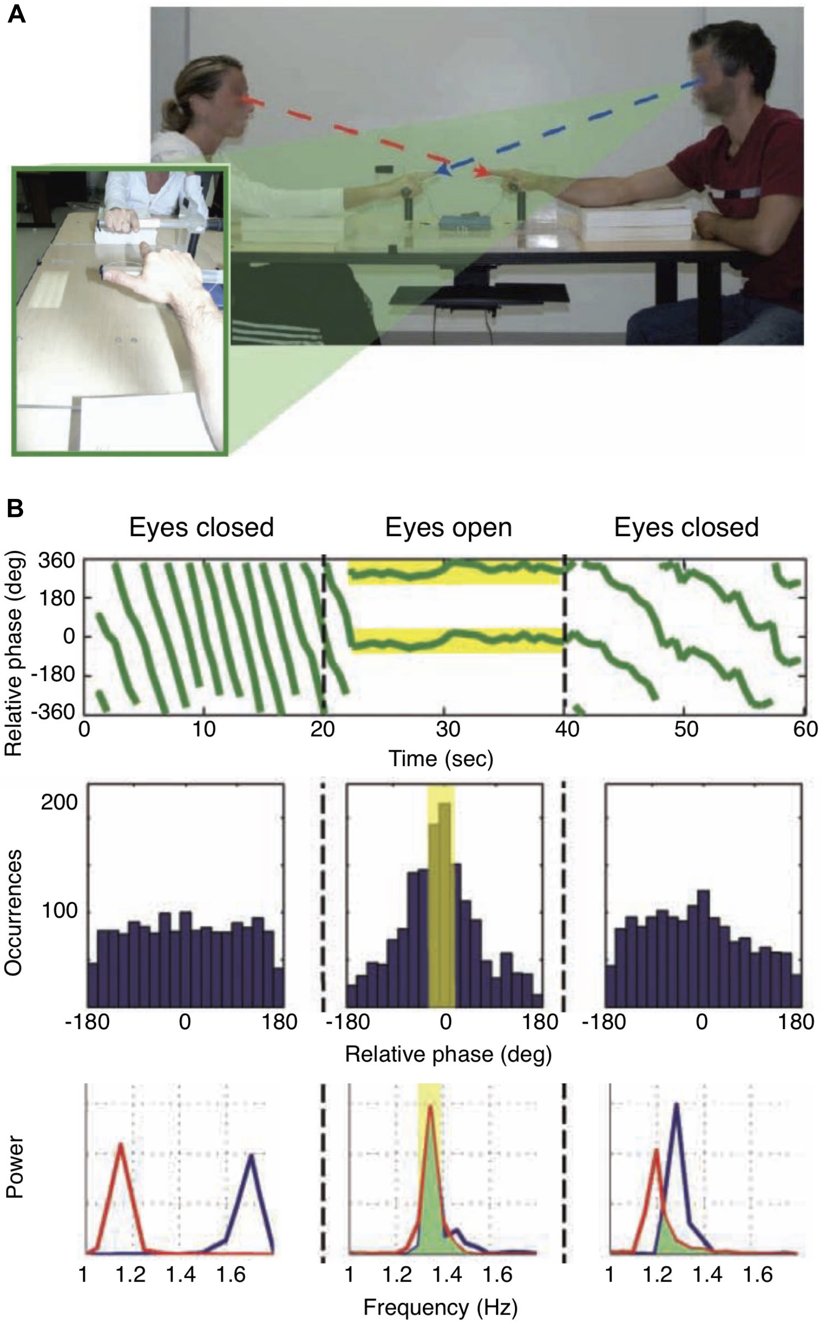


FIGURE 3 | Coordination dynamics in social interaction. **(A)** Experimental setup. Participants were seated opposite to each other and instructed to move their index finger up and down continuously, either with eyes open or eyes closed in separate periods. Importantly, no specific instructions about the coordination of the finger movements were given. **(B)** (Top) Relative phase of the finger movements, indicating synchrony when participants had their eyes open and were viewing each other's movements. (Middle) Occurrence of relative phase lags of movements. With eyes open, zero phase lag dominated the distribution. (Bottom) With eyes open, participants adopted the same movement frequency; of note, movement frequencies remained similar when participants closed their eyes again. Modified from Oullier et al. (2008).

(Walton et al., 2018). Seeking to understand how musicians communicate and engage socially in an under-determined performance context, Walton et al. (2018) ascribe a central role to shared temporal structure that provides the foundation for performers to interpret and respond to the acts of their partners. Such shared rhythms may provide the basis for what we have termed unite SMCs and for more complex forms of social expression.

It should be emphasized that coordination dynamics is, of course, also relevant in non-rhythmic behaviors. Joint attention may serve as an example here (Sebanz et al., 2006). Joint attention is an important feature of social interaction, consisting in the capability of several agents to simultaneously direct their attention toward the same object. The capacity for engaging in joint attention is frequently taken to indicate the deployment of theory of mind in the participating agents. However, the prominence of sensorimotor components in establishing and sustaining episodes of joint attention, e.g., eye and head movements, pointing and vocalizations, suggests that the concept of socSMCs may be well-placed to account for important parts of joint attention without the need to invoke theory of mind abilities (Maye et al., 2017). For example, exchanging looks or alternating gaze direction between the partner and the object of interest is a simple but powerful mechanism that can establish the mutual awareness of being jointly engaged in a perceptual episode. In addition to gaze perception, head and body orientation may be used as well to infer the target of attention. This view receives support from behavioral studies in humans showing that providing the partners with information about each other's gaze can significantly enhance performance in a collaborative search task (Wahn et al., 2015). The socSMCs concept refutes the necessity of explicitly detecting and representing the attentional state of an interaction partner. Rather, it highlights the efficacy of the co-attender in modulating the interaction between both partners and between them and the attended object. This transforms the problem of detecting a state into one of establishing a coupling. Jointly attending agents are then organized through this coupling, offering them opportunity windows of coordinated engagement (Fantasia et al., 2014). Common foci of attention are not just passively shared; rather, the co-attenders also shape them, extend them over time by embedding them in task contexts and conventionalize them in terms of canonical forms in the culture (Bruner, 1995).

Similar conclusions are suggested by developmental studies on joint attention. Humans engage in reciprocal attention from as early on as their first hour of life (Trevarthen, 2005; Reddy, 2008; Reddy and Uithol, 2016). Studying vocalizations, movement and gaze of infants interacting with their caregivers, key findings from this field of research include that infants easily follow others' gaze with their own (Hood et al., 1998; Moore and Corkum, 1998), respond meaningfully even to actions they themselves cannot produce (i.e., their capacities go beyond spectatorial mirroring) and joyfully enter into mutual responding with others, with whom they co-create rhythms and narratives. These developmental steps provide examples for the acquisition of what we term check SMCs and sync SMCs (**Figure 1**). We grow up in a field of social relations that offer opportunities to participate in joint attention settings, leading us to acquire a know-how about others as bearers of intentions (Reddy, 2003; De Jaegher et al., 2010). Thus, joint attention may be seen as an example for how sensorimotor coupling can lead to an alignment of the agents at the perceptual-motor level as a basic mechanism that contributes to mental alignment in joint action. This may be a seen as preparatory stage for the development of the capability to implicitly take another's perspective in cooperative situations and later to explicitly understand the other's perspective as such (Fuchs, 2013). We argue, furthermore, that such basic sensorimotor coordination dynamics influences, adapts and supports our more abstract ability to predict, read and engage with other's behavior and experience.

Indeed, one of the questions emerging from the socSMC concept is whether subjective feelings of social engagement are associated with motion synchronization between agents, i.e., whether the degree of social engagement can be predicted by the strength of social entrainment. To study this hypothesis one can imagine several scenarios, e.g., situations in which agents synchronize their movements, act together to achieve common goals, play music, or dance together. One study investigating this influence used a three-dimensional mirror game, in which agents had to synchronize their movements (Llobera et al., 2016). Either one of the agents was leading or following, or they jointly improvised without a designated leader and follower. The analysis of motion data and of subjective ratings revealed that the perceived sensation of synchrony could be predicted by parameters of motor synchronization in this mirror game. Especially the speed differences between the agents' movements were a good predictor for the subjective sensation of synchrony.

Several studies also used objective measures to quantify social engagement, e.g., by the duration of co-confident motion which corresponds to jitter-free, synchronous movements of two interacting agents. Co-confident motion was first described in a one-dimensional version of the mirror-game, a simple joint improvisation task (Noy et al., 2011; Hart et al., 2014; Gueugnon et al., 2016). Here, periods of co-confident motion were associated with increased social engagement and, thus, considered to indicate moments of togetherness. Even physiological parameters such as increased heart rates were shown to be associated with periods of co-confident motion and, moreover, these periods showed correlated heart rates between two improvising agents (Noy et al., 2015). We have recently obtained similar evidence in a joint attention task, in which two agents had to cooperate to determine the motion direction of a visual object on a screen. We observed that autonomic parameters related to heart rate variability could reflect the subjective evaluation of performance in the task (Maye et al., 2020). In other studies, personality traits such as the attachment style (Bowlby, 1969) were used to predict complexity and synchronization of motion in joint improvisation (Feniger-Schaal et al., 2016, 2018).

IMPAIRED SOCIAL COUPLING

The concept advocated here also has implications for understanding the basis of social cognition disorders. Impaired communication plays a role in many areas of psychiatric and psychotherapeutic practice, from temporary cases of miscommunication to persistent deviations and impaired social interactions. Communication deficits are a highly relevant aspect in diverse psychiatric disorders, such as schizophrenia and other psychotic disorders (Baltaxe and Simmons, 1995; Fioravanti et al., 2005), depression (Pope et al., 1970) and, in particular, neurodevelopmental disorders of the autism spectrum type (Magiati et al., 2014; Tillmann et al., 2019). The socSMCs concept predicts that patients with social cognitive deficits may suffer from deficits in mechanisms for interpersonal sensorimotor entrainment.

Autism spectrum disorder (ASD) may serve as a specific example for a condition with verbal as well as non-verbal communicative deviations (Lai et al., 2014). First described several decades ago in the context of schizophrenia as autistic thinking (Bleuler, 1911), autism was later investigated by Kanner (1943) and Asperger (1944) and underwent a considerable paradigm shift with the introduction of the autism spectrum (American Psychiatric Association, DSM V). Recently, ASD has been investigated extensively in the fields of psychology, psychiatry as well as clinical neuroscience (Happé and Frith, 2006; Frith and Frith, 2008; Wolfers et al., 2019). With symptoms that range from social and communicative to sensory and motor impairments, ASD's etiology and pathophysiology are still not fully understood and until today, only very few established treatment options exist.

It has been argued that reduced social entrainment in ASD may relate to impaired perception of affordances provided by other persons' behaviors (Hellendoorn, 2014). The Gibsonian notion that behavior affords behavior (Gibson, 1986) resonates well with the socSMCs concept proposed here, since it emphasizes the emergence of affordances in joint action and implies a coupling of perception-action loops supporting the social interaction (Hellendoorn, 2014). An immediate application of socSMCs principles to ASD suggests strategies for enhancing social coupling at the sensorimotor level. Brezis et al. (2017), for example, compared autistic and typically developing participants' behavior on the mirror game, an open-ended task where two players take turns leading, following, and jointly improvising motion using two handles set on parallel tracks. They found that autistic participants had lower rates and shorter duration of co-confident motion, in particular when they were following. These differences remained even when controlling for motor skills. Based on participants' subjective reports, the authors suggest attention, motivation, and reward-processing as potential mediating factors, and propose to examine the potential of specific training of sensorimotor coordination to enhance patients' social cognitive abilities. Along these lines, a recent study has investigated the impact of a dance/movement intervention on social cognition in ASD (Koehne et al., 2016). The authors observed that training of movement imitation and synchronization increased emotion inference in adults with ASD.

Another well-studied domain of impaired SMCs in ASD are eye movements. Among the most frequently observed symptoms in ASD, the avoidance of eye contact leads to a range of consequences in social interaction. Studies on human social development show that 2-year-old children with ASD tend to show significantly less visual fixation time on faces, when a video of an actress (acting as a care-giver) was presented (Jones et al., 2008), indicating a very early impairment in a social adaptive behavior that is regarded as evolutionarily vital for survival in humans and shown to be relevant for newborns at very early stages in development (Farroni et al., 2002). This early deficit seems to persist into adulthood, as shown in an eye-tracking study in adults using naturalistic social situations as stimuli (Klin et al., 2002). Importantly, this deficit also causes a lack of active perception in a critical time window in early development, in which basic learning processes drive social and emotional development, and may therefore be closely related to symptoms such as the difficulty to recognize emotional expressions in others (Eack et al., 2015). This difficulty is detrimental to any kind of communication and reported frequently in ASD as one of the most impairing symptoms. The case of gaze aversion exemplifies how active visual perception is intricately linked to both development and learning in social contexts as well as the successful unfolding of communicative acts.

Complementing these behavioral studies, neurophysiological evidence indicates that not only sensory (Robertson and Baron-Cohen, 2017) and motor (LeBarton and Landa, 2019) processing appears deviant in ASD, but also the interplay between these domains. It has been shown in children with ASD that resting state fMRI connectivity is reduced between visual and motor systems (Nebel et al., 2016). The reduction of visual-motor coupling was associated with symptom severity in terms of more severe social deficits. The socSMCs concept implies that social entrainment involves mechanisms for acquiring action-effect contingencies in the social interaction and, thus, a critical role of brain regions involved in prediction of sensory inputs and action outcomes, such as prefrontal cortex, premotor cortex, cingulate cortex, superior and middle temporal gyrus, basal ganglia and the cerebellum (Schubotz, 2007; Brown and Brüne, 2012; van der Steen and Keller, 2013; Sokolov, 2018; Van Overwalle et al., 2019). Accordingly, deficits in such predictive mechanisms should have an impact on social entrainment. Indeed, a key deficit in ASD seems to concern the ability to form flexible predictions, leading to an impairment in processing of new or unexpected sensory inputs (Gomot and Wicker, 2012) and aberrant movement planning in joint action contexts (Gonzalez et al., 2013). Deficits in predictive mechanisms in ASD have also been postulated by Sinha et al. (2014). According to their proposal, an underlying deficit in predictive abilities may account for many of the salient traits in ASD, including sensory hypersensitivities, difficulties to interact with dynamic objects, reduced motor anticipation, as well as difficulties in anticipating the actions of other persons (Sinha et al., 2014). At the neural level, this predictive impairment may relate to alterations in structures involved in prediction like the basal ganglia, anterior cingulate and cerebellum (Sinha et al., 2014; Sokolov et al., 2017; Sokolov, 2018; Van Overwalle et al., 2019). In particular, the cerebellum shows developmental alterations in ASD, including strong expression of ASD susceptibility genes, volume decreases and cellular abnormalities (Wang et al., 2014). This agrees with a role of cerebellar circuits in outcome prediction, signaling of prediction errors and perception of a person's motion and body language in social context (Sokolov et al., 2017; Sokolov, 2018; Van Overwalle et al., 2019). Deficits in sensorimotor entrainment in ASD have been examined by Wang and Hamilton (2012) and Forbes et al. (2017), who studied motor mimicry in social interaction. They observed that people with ASD can still mimic, i.e., unconsciously copy the actions of others, but do not use social cues like, e.g., gaze to control what to mimic (Forbes et al., 2017). This provides support for the hypothesis proposed here, demonstrating mimicry as a socially relevant coupling mode which influences engagement through sensorimotor entrainment.

EXTRINSIC NEURAL COUPLING MODES

To explore the neural mechanisms involved in social interaction, the concurrent observation of brain dynamics ongoing in two (or more) people who communicate, work on a joint task, or improvise together seems highly informative. In recent years, the investigation of inter-brain coupling using so-called hyperscanning methods based on simultaneous electro- or magnetoencephalographic (EEG/MEG) recordings or functional magnetic imaging (fMRI) scans of individuals engaged in a social task has gained attention in social neuroscience (Montague et al., 2002; Schippers et al., 2010; Hasson et al., 2012; Sänger et al., 2013; for a recent review also see Czeszumski et al., 2020). These approaches investigate the neural signatures of dynamic social coordination, the temporal and spatial scales on which brains interact and the correlates of behavioral coordination at the level of brain-to-brain coupling. Hyperscanning paradigms employed to investigate social interactions are manifold, including joint musical performance (Lindenberger et al., 2009; Sänger et al., 2013; Novembre et al., 2016), verbal communication (Liu et al., 2017; Li et al., 2021), decision-making in economic games (King-Casas et al., 2005; Krueger et al., 2007; Jahng et al., 2017; Hu et al., 2018), and sensorimotor coordination and imitation (Hari and Kujala, 2009; Babiloni and Astolfi, 2014; Hari et al., 2015; Liu et al., 2018; Nummenmaa et al., 2018). The intriguing idea of investigating social interactions by simultaneously recording neuronal activity from interacting brains has also been implemented for the investigation of adult-infant interactions (Hasegawa et al., 2016; Leong et al., 2017), pain perception and interpersonal touch (Goldstein et al., 2018), and has been transferred to real-life scenarios such as flight simulations in professional pilots (Toppi et al., 2016) and classroom group dynamics (Dikker et al., 2017).

To identify neural signatures of social interactions, connectivity analyses have been applied to measure both phase as well as envelope brain-to-brain coupling. The quantification of inter-brain coupling in EEG and MEG hyperscanning data includes the assessment of phase-locking between oscillatory activity in specific frequency bands (Lindenberger et al., 2009; Dumas et al., 2010; Sänger et al., 2013), as well as amplitude envelope correlations of oscillatory power (Tognoli et al., 2007; Naeem et al., 2012; Kawasaki et al., 2013). There is growing evidence from EEG/MEG hyperscanning studies that links connectivity between brains to interpersonal coordination and joint action (see for example Dumas et al., 2010; Toppi et al., 2016; Szymanski et al., 2017; Kawasaki et al., 2018; Zamm et al., 2018). Particularly, in experimental paradigms involving rhythmic, musical or motor coordination, the alpha- (or mu rhythm, oscillatory activity ranging from 8 to 13 Hz) and beta- (15–30 Hz) bands seem to mediate inter-brain coupling (Tognoli et al., 2007; Lindenberger et al., 2009; Dumas et al., 2010; Naeem et al., 2012; Novembre et al., 2016; Kawasaki et al., 2018). Besides phase relations, amplitude envelope correlations between brains are computed to investigate slower fluctuations during coordinated behavior (Hari et al., 2015; Zamm et al., 2018), which may be more appropriate considering the timescale of interpersonal sensorimotor coordination.

The socSMCs concept suggests that establishing direct links between movement kinematics and neural data recorded during social interaction might be particularly promising. One way to link neural measurements with movement data in joint action research is exemplified by the work of Zhou et al. (2016). The authors used phase-amplitude coupling to quantify the relation between the phase of hand movement accelerations and oscillatory power in the alpha- and beta-bands during a joint motor task in a dual-MEG setup (**Figure 4**). The participants had to coordinate rhythmic precision-grip-like movements while brain signals were recorded simultaneously using two MEG systems. The goal of the task was to synchronize the own movements with those of the partner, either leading or following in the interaction. The data show a movement-related modulation of alpha- and beta-band power in sensorimotor cortex and, furthermore, a modulation of beta-band power in visual cortex, which was stronger in the follower compared to the leader condition. The authors suggest that this modulation of oscillatory brain activity might be a signature of the need for the follower to coordinate own proprioceptive signals with the visual information about the movement of the leading participant (Zhou et al., 2016).

Several questions regarding the interpretation of hyperscanning results arise: (i) What is the substrate or underlying mechanism of inter-brain coupling? (ii) How can inter-brain processes shape the experience and behavior of individuals in interaction? (iii) In how far is observation at the brain-to-brain level more informative than, for example, an investigation of interpersonal sensorimotor dynamics? Given that direct coupling between neuronal ensembles of two brains can be ruled out for the lack of neuroanatomical connection, shared or synchronized sensory inputs, and coordinated motor outputs, are potential candidates. In keeping with this idea, Dumas (2011) suggested that when individuals' perception and action are coordinated, for example in a joint task, inter-brain synchrony may reflect sharing of information via between-individual sensorimotor loops or channels (Hasson and Frith, 2016; Pezzulo et al., 2019). Akin the differentiation of check, sync and unite SMCs, processes favoring the emergence of inter-brain synchrony may be described as ranging from similar external sensory stimulation of both individuals (check), reciprocal interpersonal action (sync), and group behavior that is inspired by a common ground, be it affective, informational or sociocultural (unite). Taken together, available hyperscanning studies provide evidence that sensorimotor or informational coupling between agents can be associated with inter-brain coupling of neural signals, supporting predictions that arise from the socSMCs concept.

Both phase and amplitude coupling methods have been criticized for finding spurious coupling, or hyper-connectivity non-existent in the data (Burgess, 2013; Hari et al., 2015). For example, two neuronal ensembles oscillating at the same

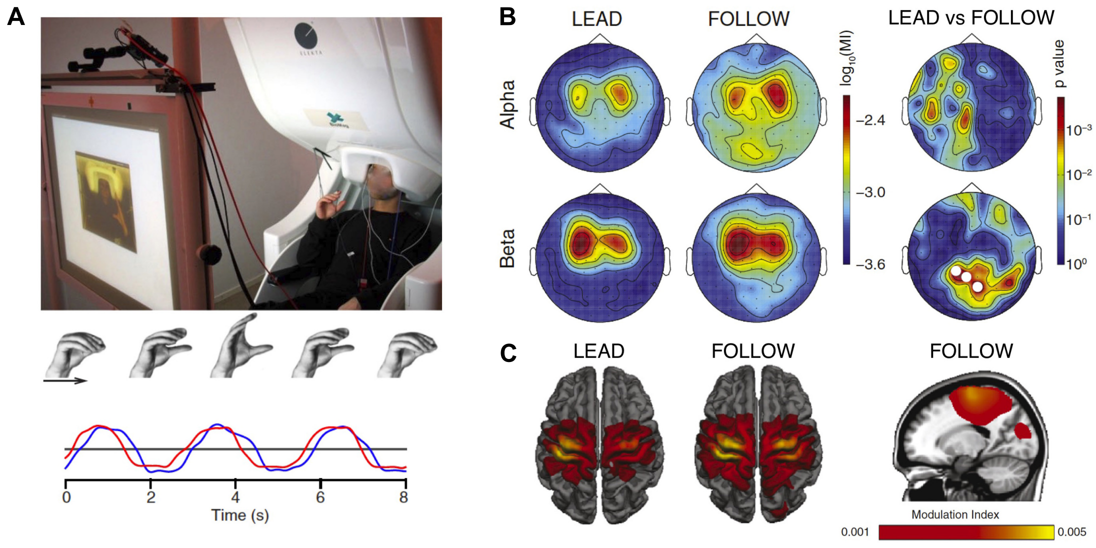


FIGURE 4 | Modulation of brain signals by joint action. **(A)** Experimental setup. Participants were seated in two separate MEG systems and instructed to perform rhythmic precision-grip-like movements in synchrony with their partner, either leading or following the other's movement. Example movement traces (red, blue) are shown at the bottom, indicating similar movement with slight delay between the participants. **(B)** Modulation of alpha- and beta-band power by the phase of the hand movement in the two conditions. Modulations occurred over central areas and, for beta power, also over visual cortex. Significant differences between the leader and follower conditions (right) occurred only for beta-band power recorded from visual areas. This role-specific modulation of brain activity might be reflecting the need for the follower to coordinate own proprioceptive signals with the visual feedback about the movement of the leading participant. **(C)** Source space projection of the results shown in panel **(B)**. Power modulations are observed in sensorimotor cortex as well as, in the follower condition, in visual cortex. Modified from Zhou et al. (2016).

frequency show high phase-locking per definition, without necessarily influencing each other. Another criticism observes that the EEG of two individuals taking part in the same experimental protocol likely shows inter-brain synchrony (due to identical sensory stimulation or similar motor output) in spite of a complete absence of interaction (Burgess, 2013; Hari et al., 2015). Circular correlation coefficients, mutual information (Burgess, 2013), or canonical correlation analyses (Campi et al., 2013; Hari et al., 2015; Vidaurre et al., 2019) have been suggested as measures that may avoid such spurious coupling. In addition, comparing inter-brain coupling in real participant pairs with randomly selected pairs (e.g., Bilek et al., 2015; Toppi et al., 2016) might aid the identification of non-trivial synchronization effects linked to the interaction between agents. However, it remains a complex task to differentiate between the diverse communicative processes involved in social interaction and to then identify their respective substrates.

The socSMCs concept argues for an integrative analysis of interaction data, including behavioral coordination in terms of sensorimotor coupling between agents, inter-brain synchronization, and subjectively experienced social engagement. A testable hypothesis is the prediction of self-assessment of social engagement, as measured by questionnaires or rating scales administered during joint action, from measures of behavioral and neural coupling between agents. Supporting this hypothesis, several studies have linked neural synchronization between interacting brains to subjective experience, e.g., feelings of engagement and social closeness (e.g., Dikker et al., 2017) or ratings of pain experience (Goldstein et al., 2018). These findings are complemented by evidence linking movement synchronization to social cohesion and subjective experience (as detailed above and also reviewed in Valencia and Froese, 2020). From the viewpoint of socSMCs, it is desirable to now go a step further and combine measures of social entrainment and social engagement, i.e., sensorimotor coupling, inter-brain synchronization and subjective experience into one model of social interaction.

Hyperscanning setups have also been used for joint neuromodulation of interacting participants, using an interventional approach to further explore underlying mechanisms of inter-brain coupling. In a study involving transcranial alternating current stimulation (tACS) applied simultaneously over motor cortex in two subjects during a joint finger tapping task, movement synchrony was enhanced by in-phase beta-band tACS (Novembre et al., 2017). Another study used dual-brain tACS to augment social interactive learning by enhancing spontaneous movement synchrony (Pan et al., 2020). Future studies might test whether such neuromodulatory interventions that lead to enhanced movement synchrony also have a potential impact on the subjects' assessment of social engagement.

As discussed earlier, we propose that the socSMCs concept might also provide new angles for neuropsychiatric research and psychological treatment, for example in ASD. Several studies have investigated interpersonally shared sensorimotor rhythms and their role for joint attention, mutual trust or empathy in hyperscanning setups involving ASD patients. These studies have revealed reduced inter-brain coupling in dyads involving

ASD participants compared to neurotypical controls, which was associated with the impairment of the social interaction and/or the severity of ASD (Tanabe et al., 2012; Salmi et al., 2013; Hasegawa et al., 2016).

RELEVANCE FOR HUMAN–ROBOT INTERACTION

We propose that the relevance of sensorimotor entrainment for social coupling not only applies to human social interaction, but can also serve to improve human–robot interaction (HRI). In fact, work in robotics provides early implementations of decentralized embodied executive control (Brooks, 1991). In the development of socSMCs-based robot controllers, the focus lies on algorithms for learning and deploying action-effect contingencies rather than for extracting semantic features from the sensor data, high-level reasoning and action planning and execution as in current mainstream robotics. The socSMCs concept suggests that many of the social action-effect contingencies involved in HRI can be observed by using rather simple features calculated from the sensory data. For example, optical flow can be used to entrain a population of neuronal oscillators by adjusting their phases and frequencies. When a motor control signal is derived from a weighted superposition of the oscillator signals, this model enables a robot to imitate gestures and to synchronize its movements with the human partner (Ansermin et al., 2016). Exploiting the mutual entrainment drastically simplifies the computational complexity of gesture mirroring and achieves millisecond-precision synchronization, which is challenging to accomplish with controllers that require high-level planning processes. Other low-level sensor data, like, for example, from distance sensors, collision detectors or the power consumption of the wheel drive, have been used to learn associations between actions and resulting changes in the sensory input, i.e., SMCs. Basically, sensor readings were combined to form an entry into a memory of SMCs that the robot had explored in the corresponding context. A reward function was used to rank different behavioral options. Together with a history of recently activated SMCs, the robot could develop an understanding of the geometric properties of its environment (Maye and Engel, 2011). This allowed the robot to traverse the space without hitting obstacles not because it was programmed to pull back whenever a distance sensor flagged an imminent collision, but because it inferred from the learned SMCs and its previous action sequence where it was and that moving on would have a detrimental effect.

The reward structure of behavioral options that is conditioned on the recent history of sensorimotor interactions can be conveniently captured by Hidden Markov Models (Maye and Engel, 2013). A powerful feature of this approach is the dual use of the model. Employed as a forward model, imagined or observed sensorimotor sequences can be used to simulate future behavioral trajectories and gauge their outcomes. In the backward direction, histories of sensorimotor interaction can be searched for common patterns which effectively is a way to derive more abstract knowledge from a set of particular interactions that all yielded the same effect.

We hypothesize that implementing social interaction capabilities in a robot which already is driven by knowledge of relevant SMCs may not depend on any critical module or function, as little as social cognition does not require any extra components that a cognitive agent wouldn't have. Therefore, adapting SMC-based robotic approaches to the social level by including socially relevant, low-level sensorimotor features seems straightforward. A model case for this transition has been made in a study which investigated a scenario where a robot and a human jointly balanced a ball on a plank (Ghadirzadeh et al., 2016). At the first stage, the robot learned the own action-effect contingencies of tilting its end of the plank and the trajectory of the ball. It then collaborated with a human by optimizing the joint goal function which kept the ball on the plank. An example for a real-world scenario that strongly relies on this type of sensorimotor coupling is the joint lifting and carrying of heavy objects, e.g., during removal of furniture to a new home. Reinforcement learning was employed for action selection from learnt SMCs, and residual uncertainty of human actions was modeled by Gaussian processes. The possibility to predict human movements from chunks of past trajectories indicates that human behavior indeed exhibits patterns which can be exploited by robot controllers (Bütepage et al., 2018). Instead of top-down approaches like explicit cost functions or target-specific training data, the authors used a bottom-up, data-driven model that was trained in an unsupervised way. Knowing regularities in the way humans move allows the controller to make predictions about the human's actions, which greatly limits the space of possible robot movement trajectories and thereby lowers response times (Bütepage et al., 2019). It has to be pointed out that this approach is different from gesture recognition in that it does not attempt to derive abstract descriptions of the movements like pointing or stirring, which is then the basis for decision making and action planning. In the socSMCs framework, the robot is rather controlled by a network of sensorimotor memory traces in which reward-based learning assigned utilities to paths and which can be used by the controller to evaluate behavioral options. More generally speaking, developing HRI on the basis of the socSMCs concept does not suggest to introduce articulated contingency detector modules. Social coordination, rather, results from linking the individual agents' networks of SMCs through the interaction, thus constituting a global network in which circular causality drives the collective dynamics. Corresponding simulation studies in evolutionary robotics have successfully modeled interaction dynamics in the perceptual crossing paradigm in which participants seek to differentiate a partner, their shadow and a static object – all of which feel the same as you cross them, only two of which move, and only one of which (the partner) responds to one's presence (Di Paolo et al., 2008).

By making human behavior more accessible for robot controllers, wearable sensors may help bridging the currently very different physical substrates of human and artificial agents and facilitate social entrainment in HRI. For example, data from a head-worn inertial measurement unit can enable a robot controller to learn human movement patterns related to mutual attentiveness, coordination and overall positivity (Hwang et al., 2019). We suggest that HRI feels natural to the extent that SMCs acquired in human-human interaction can be deployed also in the interaction with the robot. This idea has consequences for all aspects of robotic development. For example, synchronized movements, such as when we pass on or carry objects together, require mutual frequency adaptation in the human and the robot. This process runs much more efficient if the intrinsic frequency properties of the human and robotic embodiments are compatible (Ansermin et al., 2017), which can inform the mechanical design of robots, e.g., to size robotic limbs comparable to those of humans. Another effort to narrow the gap between different embodiments and make SMCs acquired in human-human interaction useful in the context of HRI may be the development of methods for endowing robots with facial expressions (Vouloutsi et al., 2019). This may be seen as a gimmick at first; however, from the socSMCs perspective, changing facial expressions support just another subset of SMCs that humans engage in their mutual interaction, which may facilitate also the interaction with the robot.

Thus, socSMCs-based human–robot coupling may enhance computational efficiency through information reduction and yield robot controllers that depend less on abstract explicit internal representations, rendering real-time control of the interaction feasible. A few iterations of the interpersonal sensorimotor loop may activate memories of previous or similar interactions which may then modulate the relative weighting of possible behavioral options that the agents can choose from. This also has the potential to replace rather discrete switching of the active role between the human and the robot with quasi-continuous turn-taking, encouraging the feeling of doing something together as opposed to interacting with a machine.

GROUNDING TOGETHERNESS IN DYNAMIC COORDINATION

As pointed out above, the socSMCs concept combines pragmatic (embodied, enactive) approaches with a constitutive role of social interaction, questioning the appropriateness of conceiving minds as independent individual entities (see also De Jaegher and Di Paolo, 2007; Gallagher, 2008; Satne and Roepstorff, 2015; Kyselo, 2016). For the study of human social capabilities, this implies a dissolution of the boundaries between me and the other that pervade classical cognitivist approaches. In particular, the socSMCs concept focuses on the relation between coupling dynamics at neurophysiological and behavioral levels, and the varying degrees of social engagement experienced by the individuals. This is in line with results from studies that used the mirror game, a simple setup in which two players sit opposite each other and coordinate the movement of two handles placed on parallel tracks in front of them. Noy et al. (2011) show that highly jitter-free, co-confident movement goes hand in hand with the highly agreeable experience of togetherness – a subjective merging of self and other, accompanied by the sense that every action is the right one. In a follow-up study, Noy et al. (2015) further showed how both subjective ratings of moving together and objective motion-based markers are predictive of physiological responses like correlated heart rate fluctuations.

The socSMCs concept also receives support from studies that highlight the role of active sensorimotor coordination for agent recognition in a simple virtual game involving perceptual crossing (Froese and Di Paolo, 2010; Auvray and Rohde, 2012; Froese et al., 2014; Lenay, 2017). In the experimental paradigm used by Auvray and Rohde (2012), two individuals move an avatar along a virtual line, on which they meet three kinds of objects: the avatar of the other player, the shadow of the other player, as well as a stationary object. While all objects feel the same (they produce a vibration) to the players, only one of them can feel and respond to co-presence: the other player's avatar. This alone suffices for players to reliably identify one another in the virtual space, based on players' ability to recognize mobile objects, as well as the fact that due to the interaction dynamic, they more frequently met their partner, versus their partner's shadow.

Another line of work that generates insight into how social engagement emerges through interaction is provided by studies of musical improvisation. For instance, Walton et al. (2018) used a combination of interviews and behavioral modeling to better understand the interactions between pairs of jazz pianists. Their models relate musicians' upper-body and musical movement (recordings of key-press timings and notes played) to changes in the musical environment (two different rhythmic background sounds), and the experience of successful and creative performance as inferred from analysis of the interviews. One of their main findings was that players' experience was heavily influenced by how well they were able to co-create a narrative – a structure to guide their collaborative play and the emergence of new behaviors. Importantly, the study demonstrates a clear relation between the movement coordination of the players and the subjective experience of social engagement, thus supporting one of the predictions of the socSMCs concept.

A closely related field of research is the study of dyadic or group improvisation in the form of dance (Himberg et al., 2018; Kimmel et al., 2018). Akin the joint creation and negotiation of time in music, Himberg et al. (2018) focus on movement coordination (quantified by motion capture) and first-person appraisal thereof (inferred from interviews and questionnaires) as a vehicle for the aesthetic experience of togetherness, i.e., moments in which dancers experience heightened connection among the group, and a genuinely distributed sense of agency. The authors establish felt togetherness as a cross-sensory and inherently shared phenomenon that clearly relates to the agents' coordination dynamics. Kimmel et al. (2018) provide a detailed phenomenological account, based on analysis of interview data, of how dancers co-create movement sequences in the explorative practice of contact improvisation. Constrained only by concerns for safety, collaboration and respect, dancers in contact improvisation deploy rolling, sliding, and falling movements to solve and create interactive challenges with their partner and the ground.

The relationship between social cohesion and interpersonal movement coordination is also revealed in experimental evidence from psychotherapeutic settings. For example, Ramseyer and Tschacher (2014) analyzed video-recorded therapy sessions and showed that both the amount of movement in patient and therapist, as well as the degree to which these movements correlate, positively predict therapeutic outcome (see also Tschacher et al., 2017; Moulder et al., 2018).

Another vast line of support for the intricate relations between bodily and personal or social dynamics comes from functional neuroanatomy. For example, the large body of work provided by Craig (2009a,b) provides detailed accounts of the neurophysiological overlap of brain regions and pathways associated with monitoring of bodily states, with areas and pathways implicated in emotion, one's subjective experience of time, and other dimensions of social and self-awareness.

Together, these findings indicate that the skill to create and express oneself in coordinative structures in real-time, together with sensitivity to one's own bodily sensations, contributes critically to the phenomenon of togetherness in social interaction dynamics. These studies support the proposal that a shared space of SMCs underlies agents' experiences of an engaging social interaction, both in the sense of being safe and predictable, as well as inviting and stimulating.

CONCLUSION

As discussed above, the socSMCs concept places joint action center stage and highlights in particular the situated and embodied sensorimotor processes that facilitate our participation in a shared social world. Our proposal, thus, extends action-oriented accounts of cognition (Varela et al., 1991; Clark, 1997; Noë, 2004; Engel et al., 2013b) to the interaction between different cognitive systems and broadens, in particular, the notion of SMCs beyond their application in the theory of individual cognition (O'Regan and Noë, 2001). In providing an overview of existing approaches to account for the complexity of dynamics present in human social cognition, we have attempted to show that novel approaches and perspectives emerge from this view of social interaction. However, key questions also remain open and need further investigation. This concerns, for instance, the exact nature of the grounding of subjective experiences of social engagement in the jointly maintained situated sensorimotor dynamics, as well as the translation of this insight into novel frameworks and interventions to support social interaction in both everyday life and clinical settings.

Pursuing the idea that SMCs may be applied in the context of social cognition, the central notion of our proposal is to ground social interaction in modes of sensorimotor and informational coupling, shifting the focus of study onto investigations of coordination dynamics as a vehicle of social entrainment. Our proposal shares aspects with interactionist concepts and joint action models of social cognition, but the socSMCs concept puts an even stronger focus on the role of low-level sensorimotor interaction dynamics for social entrainment and engagement. As we have discussed, this shift in emphasis has potential implications for the understanding of mechanisms underlying social cognition in the healthy brain but also in conditions of impaired social capabilities such as ASD. While work on the neural foundations of social cognition has, in the past decades, strongly focused on the capacity of the brain to mirror the actions of others, recent work suggests a key role for predictive mechanisms in social cognition in health and disease, and dynamic coupling between agents has become an issue of increasing interest in social neuroscience. In the context of ASD, modulation of social understanding through sensorimotor entrainment may even provide a new approach for augmentation of social capabilities. In a long-term application-oriented perspective, the socSMCs concept may also give rise to novel strategies for HRI and cooperation and may allow to introduce new concepts for robotics in training of social skills, in ambient assisted living, and caregiving.

AUTHOR CONTRIBUTIONS

AE, AM, and AL developed the core ideas of the concept discussed in this review. All authors contributed to writing of this article.

FUNDING

This work was supported by grants from the EU (project 'socSMCs,' H2020-641321) and the DFG (SFB 936/A3 and SFB TRR 169/B1/B4).

ACKNOWLEDGMENTS

We thank the other members of the socSMCs project for helpful discussions of the concept and its implications.

REFERENCES

Ansermin, E., Mostafaoui, G., Beaussé, N., and Gaussier, P. (2016). "Learning to synchronously imitate gestures using entrainment effect," in *From Animals to Animats 14. SAB 2016*, eds E. Tuci, A. Giagkos, M. Wilson, and J. Hallam (Cham: Springer), 219–231.

Ansermin, E., Mostafaoui, G., Sargentini, X., and Gaussier, P. (2017). "Unintentional entrainment effect in a context of human robot interaction: an experimental study," in *Proceedings of the 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN).* (Piscataway, NJ: IEEE), 1108–1114.

Asperger, H. (1944). Die, in Autistischen Psychopathen" im Kindesalter. *Arch. Psychiatr. Nervenkrankh.* 117, 76–136. doi: 10.1007/bf01837709

Auvray, M., and Rohde, M. (2012). Perceptual crossing: the simplest online paradigm. *Front. Hum. Neurosci.* 6:181. doi: 10.3389/fnhum.2012.00181

Babiloni, F., and Astolfi, L. (2014). Social neuroscience and hyperscanning techniques: past, present and future. *Neurosci. Biobehav. Rev.* 44, 76–93. doi: 10.1016/j.neubiorev.2012.07.006

Baltaxe, C. A., and Simmons, J. Q. III (1995). Speech and language disorders in children and adolescents with schizophrenia. *Schizophr. Bull.* 21, 677–692. doi: 10.1093/schbul/21.4.677

Bastos, A. M., and Schoffelen, J. M. (2015). A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Front. Syst. Neurosci.* 9:175. doi: 10.3389/fnsys.2015.00175

Bergström, N., Ek, C. H., Björkman, M., and Kragic, D. (2011). "Scene understanding through autonomous interactive perception," in *Proceedings of the 8th International Conference on Computer Vision Systems*, eds J. L. Crowley, B. A. Draper, and M. Thonnat (Heidelberg: Springer), 153–162.

Bilek, E., Ruf, M., Schäfer, A., Akdeniz, C., Calhoun, V. D., Schmahl, C., et al. (2015). Information flow between interacting human brains: identification, validation, and relationship to social expertise. *Proc. Natl. Acad. Sci. U.S.A.* 112, 5207–5212. doi: 10.1073/pnas.1421831112

Blakemore, S. J., and Decety, J. (2001). From the perception of action to the understanding of intention. *Nat. Rev. Neurosci.* 2, 561–567. doi: 10.1038/35086023

Bleuler, E. (ed.) (1911). "Dementia Praecox oder Gruppe der Schizophrenien," in *Handbuch der Psychiatrie*. (Wien: Franz Deuticke).

Bowlby, J. (1969). *Attachment and Loss*, Vol. 1. New York: Basic Books.

Brezis, R. S., Noy, L., Alony, T., Gotlieb, R., Cohen, R., Golland, Y., et al. (2017). Patterns of joint improvisation in adults with autism spectrum disorder. *Front. Psychol.* 8:1790. doi: 10.3389/fpsyg.2017.01790

Brooks, R. A. (1991). "Intelligence without reason," in *Proceedings of the 12th International Joint Conference on Artificial Intelligence (Sydney, Australia)*, eds R. Myopoulos and J. Reiter (San Mateo, CA: Morgan Kaufmann), 569–595.

Brown, E. C., and Brüne, M. (2012). The role of prediction in social neuroscience. *Front. Hum. Neurosci.* 6:147. doi: 10.3389/fnhum.2012.00147

Bruner, J. (1995). "From joint attention to the meeting of minds: an introduction," in *Joint Attention. Its Origins and Role in Development*, eds C. Moore and P. J. Dunham (Hove: Lawrence Erlbaum Associates), 1–14.

Burgess, A. P. (2013). On the interpretation of synchronization in EEG hyperscanning studies: a cautionary note. *Front. Hum. Neurosci.* 7:881. doi: 10.3389/fnhum.2013.00881

Bütepage, J., Kjellström, H., and Kragic, D. (2018). "Anticipating many futures: online human motion prediction and generation for human-robot interaction," in *Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA)*. (Piscataway, NJ: IEEE), 4563–4570.

Bütepage, J., Kjellström, H., and Kragic, D. (2019). "Predicting the what and how - a probabilistic semi-supervised approach to multi-task human activity modeling," in *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. (Piscataway, NJ: IEEE), 1–4.

Campi, C., Parkkonen, L. T., Hari, R., and Hyvärinen, A. (2013). Non-linear canonical correlation for joint analysis of MEG signals from two subjects. *Front. Neurosci.* 7:107. doi: 10.3389/fnins.2013.00107

Campos, J. J., Anderson, D. I., Barbu-Roth, M. A., Hubbard, E. M., Hertenstein, M. J., and Witherington, D. (2000). Travel broadens the mind. *Infancy* 1, 149–219. doi: 10.1207/S15327078IN0102_1

Carruthers, P., and Smith, P. K. (1996). *Theories of Theories of Mind*. Cambridge, MA: Cambridge Univ. Press.

Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: MIT Press.

Clark, A., and Chalmers, D. (1998). The extended mind. *Analysis* 58, 7–19.

Craig, A. D. (2009a). Emotional moments across time: a possible neural basis for time perception in the anterior insula. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 364, 1933–1942. doi: 10.1098/rstb.2009.0008

Craig, A. D. (2009b). How do you feel – now? The anterior insula and human awareness. *Nat. Rev. Neurosci.* 10, 59–70. doi: 10.1038/nrn2555

Czeszumski, A., Eustergerling, S., Lang, A., Menrath, D., Gerstenberger, M., Schuberth, S., et al. (2020). Hyperscanning: a valid method to study neural inter-brain underpinnings of social interaction. *Front. Hum. Neurosci.* 14:39. doi: 10.3389/fnhum.2020.00039

D'Ausilio, A., Novembre, G., Fadiga, L., and Keller, P. E. (2015). What can music tell us about social interaction? *Trends Cogn. Sci.* 19, 111–114. doi: 10.1016/j.tics.2015.01.005

De Jaegher, H. (2019). Loving and knowing: reflections for an engaged epistemology. *Phenom. Cogn. Sci.* [Epub ahead of print]. doi: 10.1007/s11097-019-09634-5

De Jaegher, H., and Di Paolo, E. (2007). Participatory sense-making: an enactive approach to social cognition. *Phenomenol. Cogn. Sci.* 6, 485–507. doi: 10.1007/s11097-007-9076-9

De Jaegher, H., Di Paolo, E., and Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends Cogn. Sci.* 14, 441–447. doi: 10.1016/j.tics.2010.06.009

De Jaegher, H., Pieper, B., Clénin, D., and Fuchs, T. (2017). Grasping intersubjectivity: an invitation to embody social interaction research. *Phenomenol. Cogn. Sci.* 16, 491–523. doi: 10.1007/s11097-016-9469-8

De Waal, F. B. M., and Preston, S. D. (2017). Mammalian empathy: behavioural manifestations and neural basis. *Nat. Rev. Neurosci.* 18, 498–509. doi: 10.1038/nrn.2017.72

Di Paolo, E., and De Jaegher, H. (2012). The interactive brain hypothesis. *Front. Hum. Neurosci.* 6:163. doi: 10.3389/fnhum.2012.00163

Di Paolo, E. A., Cuffari, E. C., and De Jaegher, H. (2018). *Linguistic Bodies The Continuity Between Life and Language*. Cambridge, MA: MIT Press.

Di Paolo, E. A., Rohde, M., and Iizuka, H. (2008). Sensitivity to social contingency or stability of interaction? Modelling the dynamics of perceptual crossing. *New Ideas Psych.* 26, 278–294. doi: 10.1016/j.newideapsych.2007.07.006

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J. J., et al. (2017). Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Curr. Biol.* 27, 1375–1380. doi: 10.1016/j.cub.2017.04.002

Donner, T. H., Siegel, M., Fries, P., and Engel, A. K. (2009). Buildup of choice-predictive activity in human motor cortex during perceptual decision making. *Curr. Biol.* 19, 1581–1585. doi: 10.1016/j.cub.2009.07.066

Douglas, R. J., and Martin, K. A. C. (2004). Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.* 27, 419–451. doi: 10.1146/annurev.neuro.27.070203.144152

Dumas, G. (2011). Towards a two-body neuroscience. *Commun. Integr. Biol.* 4, 349–352. doi: 10.4161/cib.4.3.15110

Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., and Garnero, L. (2010). Inter-brain synchronization during social interaction. *PLoS One* 5:e12166. doi: 10.1371/journal.pone.0012166

Durt, C., Fuchs, T., and Tewes, C. (2017). *Embodiment, Enaction, and Culture: Investigating the Constitution of the Shared World*. Cambridge, MA: MIT Press.

Eack, S. M., Mazefsky, C. A., and Minshew, N. J. (2015). Misinterpretation of facial expressions of emotion in verbal adults with Autism Spectrum Disorder. *Autism* 19, 308–315. doi: 10.1177/1362361314520755

Engel, A. K. (2010). "Directive minds: how dynamics shapes cognition," in *Enaction: Towards a New Paradigm for Cognitive Science*, eds J. Stewart, O. Gapenne, and E. Di Paolo (Cambridge, MA: MIT Press), 219–243.

Engel, A. K., Friston, K., and Kragic, D. (2016). "Introduction – Where's the action?," in *The Pragmatic Turn – Towards Action-Oriented Views in Cognitive Science. Strüngmann Forum Reports*, Vol. 18, eds A. K. Engel, K. Friston, and D. Kragic (Cambridge, MA: MIT Press), 1–15.

Engel, A. K., Gerloff, C., Hilgetag, C., and Nolte, G. (2013a). Intrinsic coupling modes: multiscale interactions in ongoing brain activity. *Neuron* 80, 867–886. doi: 10.1016/j.neuron.2013.09.038

Engel, A. K., Maye, A., Kurthen, M., and König, P. (2013b). Where's the action? The pragmatic turn in cognitive science. *Trends Cogn. Sci.* 17, 202–209. doi: 10.1016/j.tics.2013.03.006

Fantasia, V., De Jaegher, H., and Fasulo, A. (2014). We can work it out: an enactive look at cooperation. *Front. Psychol.* 5:874. doi: 10.3389/fpsyg.2014.00874

Farroni, T., Csibra, G., Simion, F., and Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proc. Natl. Acad. Sci. U.S.A.* 99, 9602–9605. doi: 10.1073/pnas.152159999

Feniger-Schaal, R., Hart, Y., Lotan, N., Koren-Karie, N., and Noy, L. (2018). The body speaks: using the mirror game to link attachment and non-verbal behavior. *Front. Psychol.* 9:1560. doi: 10.3389/fpsyg.2018.01560

Feniger-Schaal, R., Noy, L., Hart, Y., Koren-Karie, N., Mayo, A. E., and Alon, U. (2016). Would you like to play together? Adults' attachment and the mirror game. *Attach. Hum. Dev.* 18, 33–45. doi: 10.1080/14616734.2015.1109677

Fioravanti, M., Carlone, O., Vitale, B., Cinti, M. E., and Clare, L. (2005). A meta-analysis of cognitive deficits in adults with a diagnosis of schizophrenia. *Neuropsychol. Rev.* 15, 73–95. doi: 10.1007/s11065-005-6254-9

Forbes, P. A., Wang, Y., Hamilton, A. F., and de, C. (2017). STORMy interactions: gaze and the modulation of mimicry in adults on the autism spectrum. *Psychon. Bull. Rev.* 24, 529–535. doi: 10.3758/s13423-016-1136-0

Frith, C. D., Blakemore, S., and Wolpert, D. M. (2000). Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. *Brain Res. Brain Res. Rev.* 31, 357–363. doi: 10.1016/s0165-0173(99)00052-1

Frith, C. D., and Frith, U. (2008). The self and its reputation in autism. *Neuron* 57, 331–332. doi: 10.1016/j.neuron.2008.01.014

Froese, T., and Di Paolo, E. A. (2010). Modelling social interaction as perceptual crossing: an investigation into the dynamics of the interaction process. *Connect. Sci.* 22, 43–68. doi: 10.1080/09540090903197928

Froese, T., Iizuka, H., and Ikegami, T. (2014). Embodied social interaction constitutes social cognition in pairs of humans: a minimalist virtual reality experiment. *Sci. Rep.* 4:3672. doi: 10.1038/srep03672

Fuchs, T. (2013). The phenomenology and development of social perspectives. *Phenom. Cogn. Sci.* 12, 655–683. doi: 10.1007/s11097-012-9267-x

Fusaroli, R., and Tylén, K. (2016). Investigating conversational dynamics: interactive alignment, interpersonal synergy, and collective task performance. *Cogn. Sci.* 40, 145–171. doi: 10.1111/cogs.12251

Gallagher, S. (2008). Intersubjectivity in perception. *Cont. Philos. Rev.* 41, 163–178. doi: 10.1007/s11007-008-9075-8

Gallant, J. L., Connor, C. E., and Van Essen, D. C. (1998). Neural activity in areas V1, V2 and V4 during free viewing of natural scenes compared to controlled viewing. *Neuroreport* 9, 2153–2158. doi: 10.1097/00001756-199801050-00017

Gallese, V., and Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends Cogn. Sci.* 2, 493–501. doi: 10.1016/s1364-6613(98)01262-5

Gallese, V., and Lakoff, G. (2005). The brain's concepts: the role of the sensory-motor system in conceptual knowledge. *Cogn. Neuropsychol.* 22, 455–479. doi: 10.1080/02643290442000310

Gallotti, M., Fairhurst, M. T., and Frith, C. D. (2016). Alignment in social interactions. *Conscious. Cogn.* 48, 253–261. doi: 10.1016/j.concog.2016.12.002

Ghadirzadeh, A., Bütepage, J., Maki, A., Kragic, D., and Björkman, M. (2016). "A sensorimotor reinforcement learning framework for physical human-robot interaction," in *Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. (Piscataway, NJ: IEEE), 2682–2688.

Gibson, J. J. (1986). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Goldstein, P., Weissman-Fogel, I., Dumas, G., and Shamay-Tsoory, S. G. (2018). Brain-to-brain coupling during handholding is associated with pain reduction. *Proc. Natl. Acad. Sci. U.S.A.* 115, E2528–E2537. doi: 10.1073/pnas.1703643115

Gomot, M., and Wicker, B. (2012). A challenging, unpredictable world for people with autism spectrum disorder. *Int. J. Psychophysiol.* 83, 240–247. doi: 10.1016/j.ijpsycho.2011.09.017

Gonzalez, D. A., Glazebrook, C. M., Studenka, B. E., and Lyons, J. (2013). Motor interactions with another person: do individuals with autism spectrum disorder plan ahead? *Front. Integr. Neurosci.* 7:23. doi: 10.3389/fnint.2013.00023

Gueugnon, M., Salesse, R. N., Coste, A., Zhao, Z., Bardy, B. G., and Marin, L. (2016). The acquisition of socio-motor improvisation in the mirror game. *Hum. Mov. Sci.* 46, 117–128. doi: 10.1016/j.humov.2015.12.005

Happé, F., and Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *J. Autism Dev. Disord.* 36, 5–25. doi: 10.1007/s10803-005-0039-0

Hari, R., Henriksson, L., Malinen, S., and Parkkonen, L. (2015). Centrality of social interaction in human brain function. *Neuron* 88, 181–193. doi: 10.1016/j.neuron.2015.09.022

Hari, R., and Kujala, M. V. (2009). Brain basis of human social interaction: from concepts to brain imaging. *Physiol. Rev.* 89, 453–479. doi: 10.1152/physrev.00041.2007

Hart, Y., Noy, L., Feniger-Schaal, R., Mayo, A. E., and Alon, U. (2014). Individuality and togetherness in joint improvised motion. *PLoS One* 9:e87213. doi: 10.1371/journal.pone.0087213

Hasegawa, C., Ikeda, T., Yoshimura, Y., Hiraishi, H., Takahashi, T., Furutani, N., et al. (2016). Mu rhythm suppression reflects mother-child face-to-face interactions: a pilot study with simultaneous MEG recording. *Sci. Rep.* 6:34977. doi: 10.1038/srep34977

Hasson, U., and Frith, C. D. (2016). Mirroring and beyond: coupled dynamics as a generalized framework for modelling social interactions. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 371:20150366. doi: 10.1098/rstb.2015.0366

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., and Keysers, C. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends Cogn. Sci.* 16, 114–121. doi: 10.1016/j.tics.2011.12.007

Hellendoorn, A. (2014). Understanding social engagement in autism: being different in perceiving and sharing affordances. *Front. Psychol.* 5:850. doi: 10.3389/fpsyg.2014.00850

Himberg, T., Laroche, J., Bigé, R., Buchkowski, M., and Bachrach, A. (2018). Coordinated interpersonal behavior in collective dance improvisation: the aesthetics of kinaesthetic togetherness. *Behav. Sci.* 8:e23. doi: 10.3390/bs8020023

Högman, V., Björkman, M., and Kragic, D. (2013). "Interactive object classification using sensorimotor contingencies," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. (Piscataway, NJ: IEEE), 2799–2805. doi: 10.1109/IROS.2013.6696752

Hood, B. M., Willen, J. D., and Driver, J. (1998). Adult's eyes trigger shifts of visual attention in human infants. *Psychol. Sci.* 9, 131–134. doi: 10.1111/1467-9280.00024

Hu, Y., Pan, Y., Shi, X., Cai, Q., Li, X., and Cheng, X. (2018). Inter-brain synchrony and cooperation context in interactive decision making. *Biol. Psychol.* 133, 54–62. doi: 10.1016/j.biopsycho.2017.12.005

Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., et al. (2013). Dynamic functional connectivity: promises, issues, and interpretations. *Neuroimage* 80, 360–378. doi: 10.1016/j.neuroimage.2013.05.079

Hwang, T. H., Effenberg, A. O., and Blume, H. (2019). "A rapport and gait monitoring system using a single head-worn IMU during walk and talk," in *Proceedings of the 2019 IEEE International Conference on Consumer Electronics (ICCE)*. (Piscataway, NJ: IEEE), 1–5.

Jahng, J., Kralik, J. D., Hwang, D. U., and Jeong, J. (2017). Neural dynamics of two players when using nonverbal cues to gauge intentions to cooperate during the prisoner's dilemma game. *Neuroimage* 157, 263–274. doi: 10.1016/j.neuroimage.2017.06.024

Jones, W., Carr, K., and Klin, A. (2008). Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. *Arch. Gen. Psychiatry* 65, 946–954. doi: 10.1001/archpsyc.65.8.946

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child* 2, 217–250.

Kawasaki, M., Kitajo, K., and Yamaguchi, Y. (2018). Sensory-motor synchronization in the brain corresponds to behavioral synchronization between individuals. *Neuropsychologia* 119, 59–67. doi: 10.1016/j.neuropsychologia.2018.07.026

Kawasaki, M., Yamada, Y., Ushiku, Y., Miyauchi, E., and Yamaguchi, Y. (2013). Inter-brain synchronization during coordination of speech rhythm in human-to-human social interaction. *Sci. Rep.* 3:1692. doi: 10.1038/srep01692

Keller, G. B., and Mrsic-Flogel, T. D. (2018). Predictive processing: a canonical cortical computation. *Neuron* 100, 424–435. doi: 10.1016/j.neuron.2018.10.003

Keller, P. E., Novembre, G., and Hove, M. J. (2014). Rhythm in joint action: psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130394. doi: 10.1098/rstb.2013.0394

Kelso, J. A. S. (2019). "Walls and borders and strangers on the shore: on learning to live together from the perspective of the science of coordination and the complementary nature," in *Learning to Live Together: Promoting Social Harmony*, ed. J. A. S. Kelso (Cham: Springer), 77–93.

Kimmel, M., Hristova, D., and Kussmaul, K. (2018). Sources of embodied creativity: interactivity and ideation in contact improvisation. *Behav. Sci.* 8:52. doi: 10.3390/bs8060052

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., and Montague, P. R. (2005). Getting to know you: reputation and trust in a two-person economic exchange. *Science* 308, 78–83. doi: 10.1126/science.1108062

Klin, A., Jones, W., Schultz, R., Volkmar, F., and Cohen, D. (2002). Defining and quantifying the social phenotype in autism. *Am. J. Psychiatry* 159, 895–908. doi: 10.1176/appi.ajp.159.6.895

Knoblich, G., and Sebanz, N. (2008). Evolving intentions for social interaction: from entrainment to joint action. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 2021–2031. doi: 10.1098/rstb.2008.0006

Koehne, S., Behrends, A., Fairhurst, M. T., and Dziobek, I. (2016). Fostering social cognition through an imitation- and synchronization-based dance/movement intervention in adults with Autism Spectrum Disorder: a controlled proof-of-concept study. *Psychother. Psychosom.* 85, 27–35. doi: 10.1159/000441111

Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., et al. (2007). Neural correlates of trust. *Proc. Natl. Acad. Sci. U.S.A.* 104, 20084–20089. doi: 10.1073/pnas.0710103104

Kyselo, M. (2016). The minimal self needs a social update. *Phil. Psychol.* 7, 1057–1065. doi: 10.1080/09515089.2016.1214251

Lai, M. C., Lombardo, M. V., and Baron-Cohen, S. (2014). Autism. *Lancet* 383, 896–910. doi: 10.1016/S0140-6736(13)61539-1

LeBarton, E. S., and Landa, R. J. (2019). Infant motor skill predicts later expressive language and autism spectrum disorder diagnosis. *Infant. Behav. Dev.* 54, 37–47. doi: 10.1016/j.infbeh.2018.11.003

Lenay, C. (2017). Explanatory schemes for social cognition - a minimalist interaction-based approach. *Pragmatism Today* 8, 63–86.

Leong, V., Byrne, E., Clackson, K., Georgieva, S., Lam, S., and Wass, S. (2017). Speaker gaze increases information coupling between infant and adult brains. *Proc. Natl. Acad. Sci. U.S.A.* 114, 13290–13295. doi: 10.1073/pnas.1702493114

Li, Z., Li, J., Hong, B., Nolte, G., Engel, A. K., and Zhang, D. (2021). Speaker-listener neural coupling reveals an adaptive mechanism for speech comprehension in a noisy environment. *Cereb. Cortex* [Epub ahead of print]. doi: 10.1093/cercor/bhab118

Lindblom, J., and Ziemke, T. (2006). The social body in motion: cognitive development in infants and androids. *Connection Sci.* 18, 333–346. doi: 10.1080/09540090600868888

Lindenberger, U., Li, S.-C., Gruber, W., and Müller, V. (2009). Brains swinging in concert: cortical phase synchronization while playing guitar. *BMC Neurosci.* 10:22. doi: 10.1186/1471-2202-10-22

Liu, D., Liu, S., Liu, X., Zhang, C., Li, A., Jin, C., et al. (2018). Interactive brain activity: review and progress on EEG-based hyperscanning in social interactions. *Front. Psychol.* 9:1862. doi: 10.3389/fpsyg.2018.01862

Liu, Y., Piazza, E. A., Simony, E., Shewokis, P. A., Onaral, B., Hasson, U., et al. (2017). Measuring speaker-listener neural coupling with functional near infrared spectroscopy. *Sci. Rep.* 7:43293. doi: 10.1038/srep43293

Llobera, J., Charbonnier, C., Chagué, S., Preissmann, D., Antonietti, J.-P., Ansermet, F., et al. (2016). The subjective sensation of synchrony: an experimental study. *PLoS One* 11:e0147008. doi: 10.1371/journal.pone.0147008

Magiati, I., Wei Tay, X., and Howlin, P. (2014). Cognitive, language, social and behavioral outcomes in adults with autism spectrum disorders: a systematic review of longitudinal follow-up studies in adulthood. *Clin. Psychol. Rev.* 34, 73–86. doi: 10.1016/j.cpr.2013.11.002

Maravita, A., and Iriki, A. (2004). Tools for the body (schema). *Trends Cogn. Sci.* 8, 79–86. doi: 10.1016/j.tics.2003.12.008

Maye, A., and Engel, A. K. (2011). "A computational model of sensorimotor contingencies for object perception and control of behavior," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. (Piscataway, NJ: IEEE), 3810–3815.

Maye, A., and Engel, A. K. (2012). "Time scales of sensorimotor contingencies," in *Advances in Brain Inspired Cognitive Systems*, eds H. Zhang, A. Hussain, D. Liu, and Z. Wang (Heidelberg: Springer), 240–249.

Maye, A., and Engel, A. K. (2013). Extending sensorimotor contingency theory: prediction, planning, and action generation. *Adapt. Behav.* 21, 423–436. doi: 10.1177/1059712313497975

Maye, A., Isern-Mas, C., Barone, P., and Michael, J. A. (2017). Sensorimotor accounts of joint attention. *Scholarpedia* 12:42361. doi: 10.4249/scholarpedia.42361

Maye, A., Lorenz, J., Stoica, M., and Engel, A. K. (2020). Subjective evaluation of performance in a collaborative task is better predicted from autonomic response than from true achievements. *Front. Hum. Neurosci.* 14:234. doi: 10.3389/fnhum.2020.00234

McGann, M., and De Jaegher, H. (2009). Self–other contingencies: enacting social perception. *Phenomenol. Cogn. Sci.* 8, 417–437. doi: 10.1007/s11097-009-9141-7

Menary, R. A. (2010). Introduction to the special issue on 4E cognition. *Phenom. Cogn. Sci.* 9, 459–463. doi: 10.1007/s11097-010-9187-6

Merleau-Ponty, M. (1962). *Structure of Behavior*. Boston: Beacon Press.

Merleau-Ponty, M. (1963). *Phenomenology of Perception*. New York, NY: Humanities Press.

Montague, P. R., Berns, G., Cohen, J., McClure, S., Pagnoni, G., Dhamala, M., et al. (2002). Hyperscanning: simultaneous fMRI during linked social interactions. *Neuroimage* 16, 1159–1164. doi: 10.1006/nimg.2002.1150

Montani, F., Ince, R. A., Senatore, R., Arabzadeh, E., Diamond, M. E., and Panzeri, S. (2009). The impact of high-order interactions on the rate of synchronous discharge and information transmission in somatosensory cortex. *Philos. Trans. A Math. Phys. Eng. Sci.* 367, 3297–3310. doi: 10.1098/rsta.2009.0082

Moore, C., and Corkum, V. (1998). Infant gaze following based on eye direction. *Br. J. Dev. Psychol.* 16, 495–503. doi: 10.1111/j.2044-835X.1998.tb00767.x

Moore, T., Armstrong, K. M., and Fallah, M. (2003). Visuomotor origins of covert spatial attention. *Neuron* 40, 671–683. doi: 10.1016/s0896-6273(03)00716-5

Moulder, R. G., Boker, S. M., Ramseyer, F., and Tschacher, W. (2018). Determining synchrony between behavioral time series: an application of surrogate data generation for establishing falsifiable null-hypotheses. *Psychol. Methods* 23, 757–773. doi: 10.1037/met0000172

Myers, N. (2015). *Rendering Life Molecular: Models, Modelers, and Excitable Matter*. Durham, NC: Duke Univ. Press.

Myers, N., and Dumit, J. (2011). "Haptic creativity and the mid-embodiment of experimental life," in *Wiley-Blackwell Companion to the Anthropology of the Body and Embodiment*, ed. F. Mascia-Lees (Oxford: Wiley-Blackwell), 239–261.

Naeem, M., Prasad, G., Watson, D. R., and Kelso, J. S. (2012). Electrophysiological signatures of intentional social coordination in the 10-12Hz range. *Neuroimage* 59, 1795–1803. doi: 10.1016/j.neuroimage.2011.08.010

Nebel, M. B., Eloyan, A., Nettles, C. A., Sweeney, K. L., Ament, K., Ward, R. E., et al. (2016). Intrinsic visual-motor synchrony correlates with social deficits in autism. *Biol. Psychiatry* 79, 633–641. doi: 10.1016/j.biopsych.2015.08.029

Noë, A. (2004). *Action in Perception*. Cambridge, MA: MIT Press.

Novembre, G., Knoblich, G., Dunne, L., and Keller, P. E. (2017). Interpersonal synchrony enhanced through 20 Hz phase-coupled dual brain stimulation. *Soc. Cogn. Affect. Neurosci.* 12, 662–670. doi: 10.1093/scan/nsw172

Novembre, G., Sammler, D., and Keller, P. E. (2016). Neural alpha oscillations index the balance between self-other integration and segregation in real-time joint action. *Neuropsychologia* 89, 414–425. doi: 10.1016/j.neuropsychologia.2016.07.027

Noy, L., Dekel, E., and Alon, U. (2011). The mirror game as a paradigm for studying the dynamics of two people improvising motion together. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20947–20952. doi: 10.1073/pnas.1108155108

Noy, L., Levit-Binun, N., and Golland, Y. (2015). Being in the zone: physiological markers of togetherness in joint improvisation. *Front. Hum. Neurosci.* 9:187. doi: 10.3389/fnhum.2015.00187

Nummenmaa, L., Lahnakoski, J. M., and Glerean, E. (2018). Sharing the social world via intersubject neural synchronisation. *Curr. Opin. Psychol.* 24, 7–14. doi: 10.1016/j.copsyc.2018.02.021

O'Regan, J. K., and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behav. Brain Sci.* 24, 939–973. doi: 10.1017/s0140525x01000115

Oullier, O., de Guzman, G. C., Jantzen, K. J., Lagarde, J., and Kelso, J. A. S. (2008). Social coordination dynamics: measuring human bonding. *Soc. Neurosci.* 3, 178–192. doi: 10.1080/17470910701563392

Pan, Y., Novembre, G., Song, B., Zhu, Y., and Hu, Y. (2020). Dual brain stimulation enhances interpersonal learning through spontaneous movement synchrony. *Soc. Cogn. Affect. Neurosci.* 16, 210–221. doi: 10.1093/scan/nsaa080

Pezzulo, G., Donnarumma, F., Dindo, H., D'Ausilio, A., Konvalinka, I., and Castelfranchi, C. (2019). The body talks: sensorimotor communication and its brain and kinematic signatures. *Phys. Life Rev.* 28, 1–21. doi: 10.1016/j.plrev.2018.06.014

Pfeifer, R., and Bongard, J. C. (2006). *How the Body Shapes the Way We Think: A New View of Intelligence*. Cambridge, MA: MIT Press.

Pope, B., Blass, T., Siegman, A. W., and Raher, J. (1970). Anxiety and depression in speech. *J. Consult. Clin. Psychol.* 35, 128–133. doi: 10.1037/h0029659

Ramseyer, F., and Tschacher, W. (2014). Nonverbal synchrony of head- and body-movement in psychotherapy: different signals have different associations with outcome. *Front. Psychol.* 5:979. doi: 10.3389/fpsyg.2014.00979

Reddy, V. (2003). On being the object of attention: implications for self–other consciousness. *Trends Cogn. Sci.* 7, 397–402. doi: 10.1016/s1364-6613(03)00191-8

Reddy, V. (2008). *How Infants Know Minds*. Cambridge, MA: Harvard University Press.

Reddy, V., and Uithol, S. (2016). Engagement: looking beyond the mirror to understand action understanding. *Br. J. Dev. Psychol.* 34, 101–114. doi: 10.1111/bjdp.12106

Robertson, C. E., and Baron-Cohen, S. (2017). Sensory perception in autism. *Nat. Rev. Neurosci.* 18, 671–684. doi: 10.1038/nrn.2017.112

Salmi, J., Roine, U., Glerean, E., Lahnakoski, J., Nieminen-von Wendt, T., Tani, P., et al. (2013). The brains of high functioning autistic individuals do not synchronize with those of others. *Neuroimage Clin.* 3, 489–497. doi: 10.1016/j.nicl.2013.10.011

Sänger, J., Müller, V., and Lindenberger, U. (2013). Directionality in hyperbrain networks discriminates between leaders and followers in guitar duets. *Front. Hum. Neurosci.* 7:234. doi: 10.3389/fnhum.2013.00234

Satne, G., and Roepstorff, A. (2015). Introduction: from interacting agents to engaging persons. *J. Conscious. Stud.* 22, 9–23.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., et al. (2013). Toward a second-person neuroscience. *Behav. Brain Sci.* 36, 393–414. doi: 10.1017/S0140525X12000660

Schippers, M. B., Roebroeck, A., Renken, R., Nanetti, L., and Keysers, C. (2010). Mapping the information flow from one brain to another during gestural communication. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9388–9393. doi: 10.1073/pnas.1001791107

Schubotz, R. (2007). Prediction of external events with our motor system: towards a new framework. *Trends Cogn. Sci.* 11, 211–218. doi: 10.1016/j.tics.2007.02.006

Sebanz, N., Bekkering, H., and Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends Cogn. Sci.* 10, 70–76. doi: 10.1016/j.tics.2005.12.009

Sheets-Johnstone, M. (2011). *The Primacy of Movement*. Amsterdam: John Benjamins Pub.

Shimazaki, H., Amari, S., Brown, E. N., and Grün, S. (2012). State-space analysis of time-varying higher-order spike correlation for multiple neural spike train data. *PLoS Comput. Biol.* 8:e1002385. doi: 10.1371/journal.pcbi.1002385

Siegel, M., Donner, T. H., and Engel, A. K. (2012). Spectral fingerprints of large-scale neuronal interactions. *Nat. Rev. Neurosci.* 13, 121–134. doi: 10.1038/nrn3137

Sinha, P., Kjelgaard, M. M., Gandhi, T. K., Tsourides, K., Cardinaux, A. L., Pantazis, D., et al. (2014). Autism as a disorder of prediction. *Proc. Natl. Acad. Sci. U.S.A.* 111, 15220–15225. doi: 10.1073/pnas.1416797111

Sokolov, A. A. (2018). The cerebellum in social cognition. *Front. Cell. Neurosci.* 12:145. doi: 10.3389/fncel.2018.00145

Sokolov, A. A., Miall, R. C., and Ivry, R. B. (2017). The cerebellum: adaptive prediction for movement and cognition. *Trends Cogn. Sci.* 21, 313–332. doi: 10.1016/j.tics.2017.02.005

Stanley, D. A., and Adolphs, R. (2013). Toward a neural basis for social behavior. *Neuron* 80, 816–826. doi: 10.1016/j.neuron.2013.10.038

Szymanski, C., Pesquita, A., Brennan, A. A., Perdikis, D., Enns, J. T., Brick, T. R., et al. (2017). Teams on the same wavelength perform better: inter-brain phase synchronization constitutes a neural substrate for social facilitation. *Neuroimage* 152, 425–436. doi: 10.1016/j.neuroimage.2017.03.013

Tanabe, H. C., Kosaka, H., Saito, D. N., Koike, T., Hayashi, M. J., Izuma, K., et al. (2012). Hard to "tune in": neural mechanisms of live face-to-face interaction with high-functioning autistic spectrum disorder. *Front. Hum. Neurosci.* 6:268. doi: 10.3389/fnhum.2012.00268

Tillmann, J., Cáceres, A. S. J., Chatham, C. H., Crawley, D., Holt, R., Oakley, B., et al. (2019). Investigating the factors underlying adaptive functioning in autism in the EU-AIMS Longitudinal European Autism Project. *Autism Res.* 12, 645–657. doi: 10.1002/aur.2081

Tognoli, E., and Kelso, J. (2014). The metastable brain. *Neuron* 81, 35–48. doi: 10.1016/j.neuron.2013.12.022

Tognoli, E., Lagarde, J., DeGuzman, G. C., and Kelso, J. A. S. (2007). The phi complex as a neuromarker of human social coordination. *Proc. Natl. Acad. Sci. U.S.A.* 104, 8190–8195. doi: 10.1073/pnas.0611453104

Tognoli, E., Zhang, M., Fuchs, A., Beetle, C., and Kelso, J. A. S. (2020). Coordination dynamics: a foundation for understanding social behavior. *Front. Hum. Neurosci.* 14:317. doi: 10.3389/fnhum.2020.00317

Toppi, J., Borghini, G., Petti, M., He, E. J., De Giusti, V., He, B., et al. (2016). Investigating cooperative behavior in ecological settings: an EEG hyperscanning study. *PLoS One* 11:e0154236. doi: 10.1371/journal.pone.0154236

Trevarthen, C. (2005). First things first: infants make good use of the sympathetic rhythm of imitation, without reason or language. *J. Child Psychother.* 31, 91–113. doi: 10.1080/00754170500079651

Tschacher, W., Giersch, F., and Friston, K. (2017). Embodiment and schizophrenia: a review of implications and applications. *Schizophr. Bull.* 43, 745–753. doi: 10.1093/schbul/sbw220

Valencia, A. L., and Froese, T. (2020). What binds us? Inter-brain neural synchronization and its implications for theories of human consciousness. *Neurosci. Conscious.* 2020:niaa010. doi: 10.1093/nc/niaa010

van der Steen, M. C. M., and Keller, P. E. (2013). The ADaptation and Anticipation Model (ADAM) of sensorimotor synchronization. *Front. Hum. Neurosci.* 7:253. doi: 10.3389/fnhum.2013.00253

Van Overwalle, F., Manto, M., Leggio, M., and Delgado-García, J. M. (2019). The sequencing process generated by the cerebellum crucially contributes to social interactions. *Med. Hypotheses.* 128, 33–42. doi: 10.1016/j.mehy.2019.05.014

Varela, F. J., Thompson, E., and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.

Vidaurre, C., Nolte, G., de Vries, I. E. J., Gómez, M., Boonstra, T. W., Müller, K. R., et al. (2019). Canonical maximization of coherence: a novel tool for investigation of neuronal interactions between two datasets. *Neuroimage* 201:116009. doi: 10.1016/j.neuroimage.2019.116009

Vouloutsi, V., Grechuta, K., and Verschure, P. F. (2019). "Evaluation of the facial expressions of a humanoid robot," in *Proceedings of the Conference on Biomimetic and Biohybrid Systems. Living Machines 2019*, ed. U. Martinez-Hernandez (Cham: Springer), 365–368.

Wahn, B., Schwandt, J., Krüger, M., Crafa, D., Nunnendorf, V., and König, P. (2015). Multisensory teamwork: using a tactile or an auditory display to exchange gaze information improves performance in joint visual search. *Ergonomics* 59, 781–791. doi: 10.1080/00140139.2015.1099742

Walton, A., Washburn, A., Langland-Hassan, P., Chemero, A., Kloos, H., and Richardson, M. J. (2018). Creating time: social collaboration in music improvisation. *Topics Cogn. Sci.* 10, 95–119. doi: 10.1111/tops.12306

Wang, S. S.-H., Kloth, A. D., and Badura, A. (2014). The cerebellum, sensitive periods, and autism. *Neuron* 83, 518–532. doi: 10.1016/j.neuron.2014.07.016

Wang, Y., and Hamilton, A. F. D. C. (2012). Social top-down response modulation (STORM): a model of the control of mimicry in social interaction. *Front. Hum. Neurosci.* 6:153. doi: 10.3389/fnhum.2012.00153

Wolf, T., Vesper, C., Sebanz, N., Keller, P. E., and Knoblich, G. (2019). Combining phase advancement and period correction explains rushing during joint rhythmic activities. *Sci. Rep.* 9:9350. doi: 10.1038/s41598-019-45601-5

Wolfers, T., Floris, D. L., Dinga, R., Rooij, D., Isakoglou, C., Kia, S. M., et al. (2019). From pattern classification to stratification: towards conceptualizing the heterogeneity of Autism Spectrum Disorder. *Neurosci. Biobehav. Rev.* 104, 240–254. doi: 10.1016/j.neubiorev.2019.07.010

Zamm, A., Debener, S., Bauer, A., Bleichner, M. G., Demos, A. P., and Palmer, C. (2018). Amplitude envelope correlations measure synchronous cortical oscillations in performing musicians. *Ann. N. Y. Acad. Sci.* 1423, 251–263. doi: 10.1111/nyas.13738

Zhou, G., Bourguignon, M., Parkkonen, L., and Hari, R. (2016). Neural signatures of hand kinematics in leaders vs. followers: a dual-MEG study. *Neuroimage* 125, 731–738. doi: 10.1016/j.neuroimage.2015.11.002

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Human Neuroscience

ORIGINAL RESEARCH
published: 05 November 2021
doi: 10.3389/fnhum.2021.767208

Neuronal Oscillatory Signatures of Joint Attention and Intersubjectivity in Arrhythmic Coaction

Alexander Maÿe*, Tiezhi Wang and Andreas K. Engel

Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

Hyper-brain studies analyze the brain activity of two or more individuals during some form of interaction. Several studies found signs of inter-subject brain activity coordination, such as power and phase synchronization or information flow. This hyper-brain coordination is frequently studied in paradigms which induce rhythms or even synchronization, e.g., by mirroring movements, turn-based activity in card or economic games, or joint music making. It is therefore interesting to figure out in how far coordinated brain activity may be induced by a rhythmicity in the task and/or the sensory feedback that the partners receive. We therefore studied the EEG brain activity of dyads in a task that required the smooth pursuit of a target and did not involve any extrinsic rhythms. Partners controlled orthogonal axes of the two-dimensional motion of an object that had to be kept on the target. Using several methods for analyzing hyper-brain coupling, we could not detect signs of coordinated brain activity. However, we found several brain regions in which the frequency-specific activity significantly correlated with the objective task performance, the subjective experience thereof, and of the collaboration. Activity in these regions has been linked to motor control, sensorimotor integration, executive control and emotional processing. Our results suggest that neural correlates of intersubjectivity encompass large parts of brain areas that are considered to be involved in sensorimotor control without necessarily coordinating their activity across agents.

Keywords: non-rhythmic interaction, self-perception, joint attention, social cognition, hyper-scanning, EEG

Edited by:
Tetsuo Kida,
Institute for Developmental Research,
Japan

Reviewed by:
Jelmer Pieter Borst,
University of Groningen, Netherlands
Olga Mikhailovna Bazanova,
State Research Institute of
Neuroscience and Medicine, Russia

****Correspondence:***
Alexander Maÿe
a.maye@uke.de

Specialty section:
This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

Received: *30 August 2021*
Accepted: *11 October 2021*
Published: *05 November 2021*

Citation:
Maÿe A, Wang T and Engel AK (2021)
Neuronal Oscillatory Signatures of
Joint Attention and Intersubjectivity in
Arrhythmic Coaction.
Front. Hum. Neurosci. 15:767208.
doi: 10.3389/fnhum.2021.767208

1. INTRODUCTION

Hyper-scanning is a term that describes the simultaneous recording of brain activity from two or more people while they undergo some form of interaction, and it has developed into an important empirical method for research in social cognition. The approach can employ all major signal modalities that are used in brain research, i.e., EEG, MEG, fMRI, and fNIRS. A consistent finding in hyper-scanning studies is that the brain activities of the interacting partners are temporally coordinated. Depending on the kind of the interaction, this coordination has been observed on different timescales and in different brain areas, suggesting that it involves a multitude of cognitive functions.

There is a spectrum of opinions about what functional role coordinated inter-brain activity might play (Hamilton, 2020). As discussed in detail by Konvalinka and Roepstorff (2012), hyper-brain activity is considered as substrate for a functional coupling between individuals that is used to underpin representational concepts of social cognition as well as dynamic or enactive accounts. In representational concepts, inter-brain coupling is seen as an enabling mechanism for shared representations of the task, goals, actions and intentions (Sebanz et al., 2006; Vesper and Sebanz, 2016). Enactive accounts emphasize the importance of the circular dynamics between interacting agents for social cognition (Fuchs and Jaegher, 2009; De Jaegher et al., 2010). More recently it has been suggested that social cognition at least in part relies on action-effect contingencies that agents deploy to predict their own and other's actions (Maye and Engel, 2016). In this framework, hyper-brain activity would be seen as an indicator for the dynamical informational and sensorimotor coupling of agents in the interaction (Lübbert et al., 2021).

The paradigms that are used in hyper-scanning studies can be roughly separated into those in which the interaction happens in turns, like in card games, or is continuous, like imitating movements. The prisoner's dilemma game may serve as a representative example for a turn-based interaction paradigm. In each turn, the players decide whether to "cooperate," "defect" or adopt a "tit-for-tat" strategy, and the combination of responses determines the reward received by the dyad. By quantifying information flow between the two brains during the period when partners make their decision by partial directed coherence, networks have been observed that change their topology depending on the combination of strategies the partners followed (Babiloni et al., 2007a; De Vico Fallani et al., 2010). An example paradigm for studying inter-brain coupling during continuous interaction is making music together. Guitarists showed increased inter-brain phase synchronization in the theta band at fronto-central electrodes when setting their tempo to the beats of a metronome and around the onset of playing a short melody together (Lindenberger et al., 2009; Sänger et al., 2012). After the onset however, this synchronization decreased. Since the reported inter-brain synchronization effects were all in the low-frequency bands, the authors suggest they may result from the similarity of the temporal properties of sensorimotor processes in the individuals.

One problem in hyper-scanning studies is to disentangle coherent inter-brain activity related to the interaction and merely correlated activity resulting from common input (Hamilton, 2020). Such correlated activity has been demonstrated between participants watching the same movie but not at the same time (Hasson et al., 2004), so that there clearly is no interaction between them. Another issue, in our view, is that the majority of tasks in hyper-scanning studies requires that the physical activity of the partners be temporally coordinated; therefore, the question arises in how far the observed inter-brain coupling can be traced back to the temporal coordination required by the task structure. In the aforementioned guitar players study, the authors suggest that the observed coupling of brain activity in the low frequency range is likely related to the partners coordinating their behavior through reciprocal sensorimotor feedback. Konvalinka and Roepstorff (2012) must have had a similar feeling when they wrote:

> "Therefore, as mutual interaction involves behavioral coupling between two people producing similar actions, and engages similar cognitive processes (such as predicting each other's actions, imitating each other's hand/finger movements, and jointly attending to joint actions) between interacting partners, it may not be so surprising that their brain rhythms are synchronized." (p. 7)

This begs the obvious question of how much and what kind of inter-brain coupling one would observe if the task would not impose strong temporal coordination of behavior. Corresponding studies are scarce though. The room cleaning scenario described in Dodel et al. (2010) may be one of the few examples. The task for the team is to enter a previously unseen room and "clean" it by keeping enemies who are potentially lingering in the room in check. Solving the task requires the team members coordinating their behavior on the tactical level without involving precise synchronization of the actions. In a hyper-scanning version of the paradigm in a virtual environment, the researchers found changes in the intrinsic dimensionality of brain activity with exercising the task, but they did not report any inter-brain coupling results (Dodel et al., 2011). Another study that observed the unfolding of coordinated room "cleaning" in a simulated environment (Tognoli et al., 2011a) reported candidate neuromarkers comprised of different EEG topographies and different frequency bands that characterize events during the action, but it also did not report any inter-brain coupling (Tognoli et al., 2011b).

To help fill this gap we investigated coupling of neuronal oscillatory activity in a joint target-tracking task. In this game of skill, two players had to adjust two orthogonal axes of a tablet in order to make a virtual ball roll toward a moving target and follow it as closely as possible. Each player minimized the distance between the ball and the target along his or her axis of control, but whether they hit the target at the same time or not was not relevant for the success. We show that the task induced oscillatory activity in sensory as well as motor areas of the brain, which is an obvious expectation if one looks for neuronal mechanisms of social interaction. We go one step further though and demonstrate that these activity patterns co-vary with the subjective experience of the interaction in terms of the own performance in the task as well as the success of the collaboration. We were particularly interested to figure out in how far properties of the inter-brain coupling would be related to the subjective experience. To address the problem of common input, we contrasted two conditions in which both players received the same input but worked on the task together or individually.

This article complements a previous report about the same dataset in which we analyzed physiological signals (heart beat, respiration, skin conductance). The analyses revealed that autonomic parameters during the game are predictive for the self-assessment of the own performance and the success of the collaboration to be rated after each trial (Maye et al., 2020). Taking this finding as an indication for the role of bodily processes in the emergence of intersubjectivity, we here aim to elucidate the contribution of activity in the brain to this experience.

2. METHODS

2.1. Joint Target Tracking Task

The main objective in the search for an experimental paradigm was that it should make the partners continuously collaborate on a task but without the need for rhythmic temporal coordination. In order to be optimally sensitive for identifying underlying EEG dynamics of implicit interpersonal interaction, the second objective was to minimize artifacts induced by movements of the subjects.

We therefore decided to adopt a virtualized version of the common labyrinth game in which one or more players move a ball to a target location by tilting the game board. In order to keep the two players in continuous interaction, we employed a moving target that reversed its direction at random intervals (see **Figure 1A**). Players controlled the tilt angle of the board along orthogonal axes. This non-redundant control mode did not allow one player to compensate errors made by the other; instead, both players independently minimized the distance between the ball and the target along their control axis. The maximum distance the target moved in one direction was 14 cm, and it traveled at about 0.86 cm/s.

Dyads were instructed to keep the ball as precisely as possible on the target. We used the Euclidean distance d between the ball and the target, accumulated along the duration of a round (T), as an objective measure of performance:

$$d = \sum_{t=1}^{T} \sqrt{\Delta x^2 + \Delta y^2},$$

where Δx and Δy are the distances along each player's axis.

We contrasted this collaborative condition with another configuration which was similar in its sensorimotor feedback but different in the level of interaction. In the individual condition, there were two balls and two targets, and each player tried to bring them together individually (see **Figure 1B**). Both targets still reversed their movement direction at random times; however, they did so at the same time. The reversal times were different in each round, thus minimizing the temporal coherence of common input across trials and dyads.

The paradigm was implemented on a tablet computer (iPad2, Apple Inc.). The kinematic of the virtual ball was driven by Newton's second law with the accelerations given by the tilt angle of the tablet. The tablet was mounted on a ball joint that carried its weight. Participants poked their index finger into a thimble on two sides, clenching the other fingers around a handle and resting their forearm on an L-shaped frame (see **Figure 1C**). This arrangement enabled the players to tilt the tablet by effortless, miniature movements of the index finger, minimizing artifacts in the EEG caused by muscle contractions.

2.2. Participants and Experimental Procedure

Twenty-eight subjects participated in the study (20 females, mean age 25.18 ± 3.86 years). All participants were right-handed and reported to be in healthy condition. Except for 2 dyads, all participants declared to not know each other in the first session. We obtained written informed consent before commencing the experiment, and the participants received financial compensation. The study was approved by the ethics committee of the medical association of the city of Hamburg.

Participants exercised the task for 6 consecutive days with the same partner. On each day, they completed 7 rounds in each condition (collaborative/individual) in random order. Each round lasted about 2 min. The training phase allowed the participants to develop a dynamically stable performance on the task and to acquire a routine for the experimental procedure. Hyper-scanning was carried out on days 7 and 8. On day 7, they played with their training partner; on day 8, they played with another subject from the cohort. The analysis did not reveal any significant differences between the data from days 7 and 8; therefore, and to support the statistical power of the analyses presented below, data from both days were pooled.

Immediately after each trial, the experimenter requested the participants to rate their feeling regarding the following aspects:

R1 : "Please rate your own performance."
R2 : "Please rate your partner's performance."
R3 : "Please rate the collaboration."

Participants selected a number between 1 and 9 (1-very poor, 9-excellent) on a small remote control which they held in their left hands underneath the armrest so that the partner could not see their selection. R2 and R3 were called out only after a collaborative trial. We consider responses to R1 and R2 as subjective measures of performance, whereas R3 reflected the success as a team.

2.3. Data Recording and Analysis

The experiment took place in an electromagnetically shielded and sound attenuated chamber. EEG was recorded simultaneously from 64 active electrodes on the scalp of each participant using two synchronized amplifiers (ActiveTwo AD-box, BioSemi instrumentation) with a sample rate of 2,048 Hz. Electrodes were placed according to the international 10/20 system and mounted in stock head caps from the same company. Electrode montage was standardized by centering Cz between nasion and inion and between the pre-auricular points.

Horizontal eye movements were recorded from two electrodes placed at the outer canthi. To record vertical eye movements, two more electrodes were mounted above and below the right eye of each participant (see subject on the left in **Figure 1C**). The horizontal and vertical components of the EOG were determined by subtracting the signals from the corresponding electrodes.

We also recorded physiological signals, i.e., ECG, respiration, EDA, and finger EMG. Details about the recording of those data and their analysis are given in Maye et al. (2020).

Data were analyzed in Matlab (The Mathworks, Natick, MA, USA) using the Fieldtrip toolbox (Oostenveld et al., 2011).

2.3.1. Preprocessing

The recorded EEG data were segmented into epochs of 109.7 s aligned to the start of each round. A 0.5 s zero-padding was added at the beginning and the end respectively. The data were

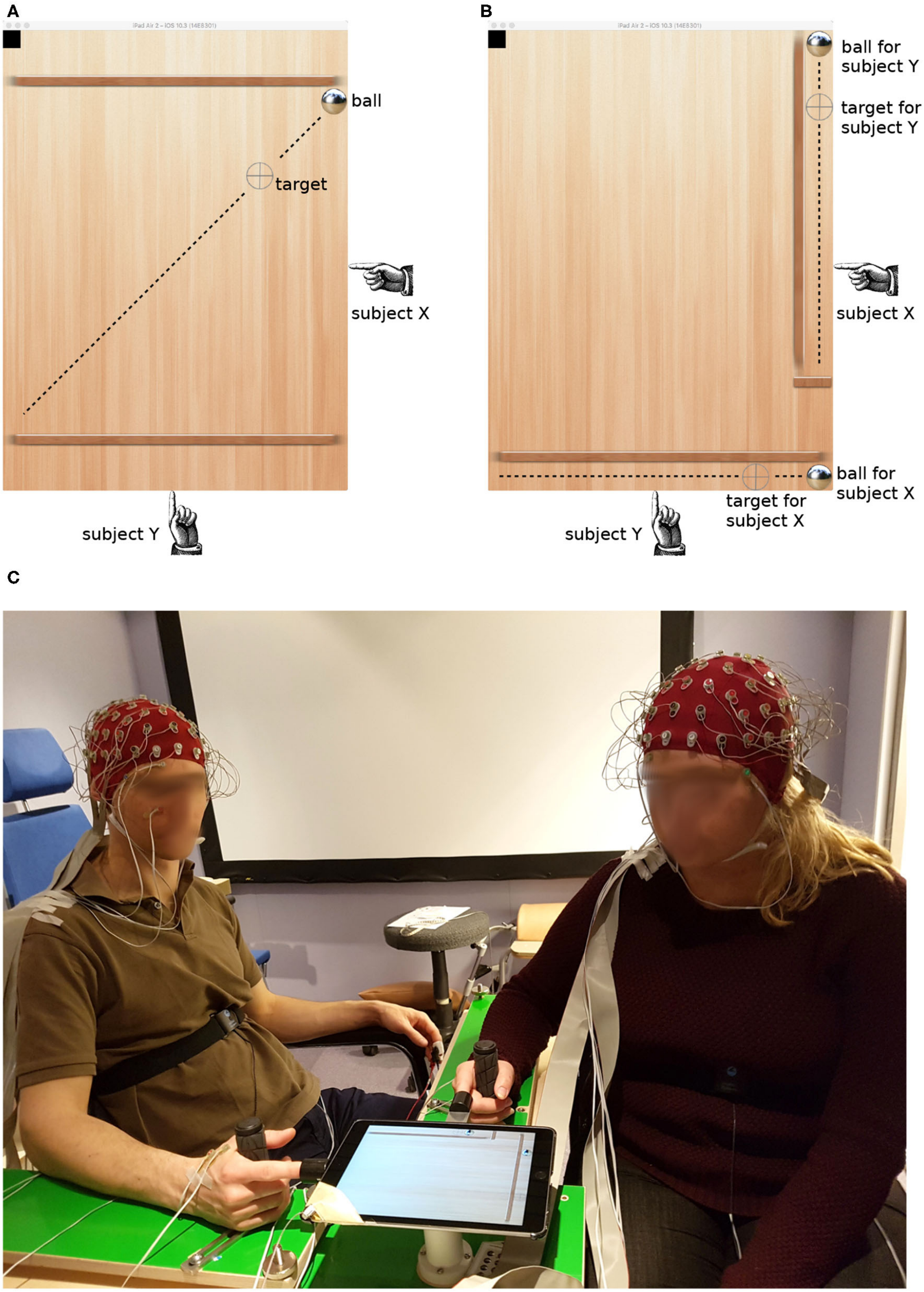


FIGURE 1 | Target tracking task in the collaborative **(A)** and individual condition **(B)**. Dashed lines visualize the trajectory of the target; they were not visible to the subjects. The picture **(C)** shows the experimental setup and two participants with EEG caps.

then re-referenced to the common average and filtered by a two-pass finite impulse response filter with Hamming window in the frequency band of 0.5–256 Hz. A notch filter was used to remove power line noise and its harmonics. Linear trends in the EEG data were also removed. The data were then resampled to 512 Hz.

In order to remove artifacts resulting from muscle activity, eye movements and blinks from the EEG data, we employed independent component analysis (ICA) over all 64 EEG channels. The artifact-typical components were manually identified, verified and removed. Artifactual components were identified by matching their patterns and time courses to those shown in Jung et al. (2000). In particular, artifacts resulting from eye movements and blinks, muscular activity on the scalp and in the neck and from cardiac activity were detected and removed. The EOG was not included in the ICA but was used for verifying the correctness of ICA. The time course of components with typical eye movement-related patterns were compared with the EOG to make sure all ocular artifact-related peaks were tracked by the ICA components with the typical patterns.

2.3.2. Localizing EEG Sources

A three-dimensional current distribution was reconstructed from the signals of the 64 scalp electrodes by eLORETA (Low Resolution Electromagnetic Tomography, Pascual-Marqui et al. 1994). This method offers exact, zero error localization to point-test sources (Pascual-Marqui, 2007). The standard boundary element method (BEM) head model provided by the Fieldtrip toolbox was used (Oostenveld et al., 2003) with a voxel edge length of 1 cm. EEG data were filtered by the inverse solution, yielding an activity trace at each voxel. The spectra of these source activities were calculated by multi-taper FFT and correlated with the behavioral indicators using Pearson's coefficient of correlation.

2.3.3. Coupling Analysis

Data were divided into overlapping windows of 2 s and 1.5 s overlap. Using 1-s-windows did not qualitatively change the results. Since our analyses are focused on oscillatory brain activity, we used multi-taper FFT to transform EEG data to the time-frequency domain. Tapers were calculated from discrete prolate spheroidal sequences (DPSS), and 3 tapers were used for calculating the complex spectrum for each data window (1,024 samples). The frequency range was 1–120 Hz in steps of 1 Hz.

Rather than calculating phase coherency across trials, we here were interested in the stability of phase differences across time, i.e., the duration of a game. To this end we calculated the auto-cross-spectra S_X and S_Y, the hyper-cross-spectrum S_{XY}, and calculated coherence C by:

$$C = \frac{|S_{XY}|^2}{S_X S_Y}$$

Using circular coherence (Burgess, 2013) did not qualitatively change the results.

We also calculated other hyper-brain coupling measures that are frequently used in the literature: amplitude or power correlation (AC/PC), phase-locking value (PLV, Lachaux et al., 1999), partial directed coherence (PDC, Baccalá and Sameshima, 2001; Baccalá et al., 2007), and directed transfer function (DTF, Kaminski and Blinowska, 1991). We used the standardized interface to these methods that is provided by the Fieldtrip toolbox.

2.3.4. Statistical Evaluation

The main comparison is between the strength of the respective coupling measure in the collaborative and the individual condition. Contrasting coupling at the sensor level and in source space likewise involves a combinatorially large number of comparisons (Maris et al., 2007). We employed cluster-based permutation tests to counteract the multiple comparison problem. This non-parametric test has the capacity for incorporating biophysical constraints (Maris and Oostenveld, 2007), which here is that electrodes/voxels with differences in coupling between the conditions be compact and that similar differences be present also in nearby frequency bands. The main idea is to compare a test statistic for the condition difference with the distribution of the statistic when the comparison is made between data that have been randomly sampled from both conditions. We used 1,000 repetitions for this resampling process. For the test statistic, we employed a dependent samples t-test with a threshold of 0.05. As we did not have a hypothesis about the strength of coupling in either condition, we considered both tails in the permutation test.

In order to relate activity clusters to the literature, locations of maximum correlation were looked up in the brainnetome atlas[1] (Fan et al., 2016) and the neurosynth database.[2]

3. RESULTS

3.1. Correlations Between Ratings

A previous analysis of the behavioral data revealed that there was no relation in how the two partners in a dyad evaluated their task performance with respect to the questions R1–R3. However, the three ratings were significantly correlated within individuals (Maye et al., 2020). Here we summarize this finding by showing the distribution of intra- and inter-individual correlations of the ratings in **Figure 2**. Whereas individual ratings were correlated with a coefficient of 0.6 or larger on average, the average inter-individual correlation of responses was around zero. Only when rating the partner's performance (R2), responses were correlated with a coefficient of about 0.4. The relation between the self-evaluation of the own performance (R1) and the tracking error (d) as an objective performance indicator was surprisingly weak (median correlation: −0.33), suggesting that participants did not accurately reflect upon their actual task performance.

3.2. Power Differences

We calculated the power spectrum for each trial and compared it between the two conditions. Across all participants, power in the range from 66 to 120 Hz was larger in the collaborative condition

---

[1] atlas.brainnetome.org

[2] neurosynth.org

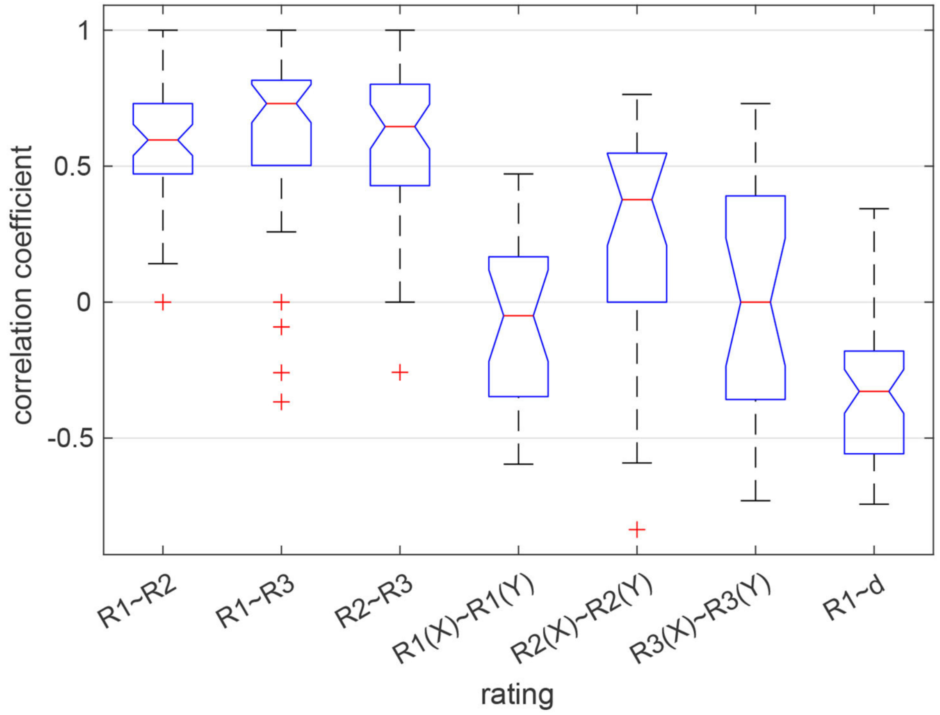


FIGURE 2 | Correlation between the ratings given by each participant (3 boxes on the left) and between the partners in a dyad (next 3 boxes). The box on the right shows the distribution of correlations between the self-assessed performance (R1) and the time-accumulated distance between the ball and the target (d). Red lines show the median, blue boxes the 25 and 75% quantiles, whiskers the most extreme values not considered outliers, and red crosses show outliers. Medians are different at the 5% level if the notches do not overlap.

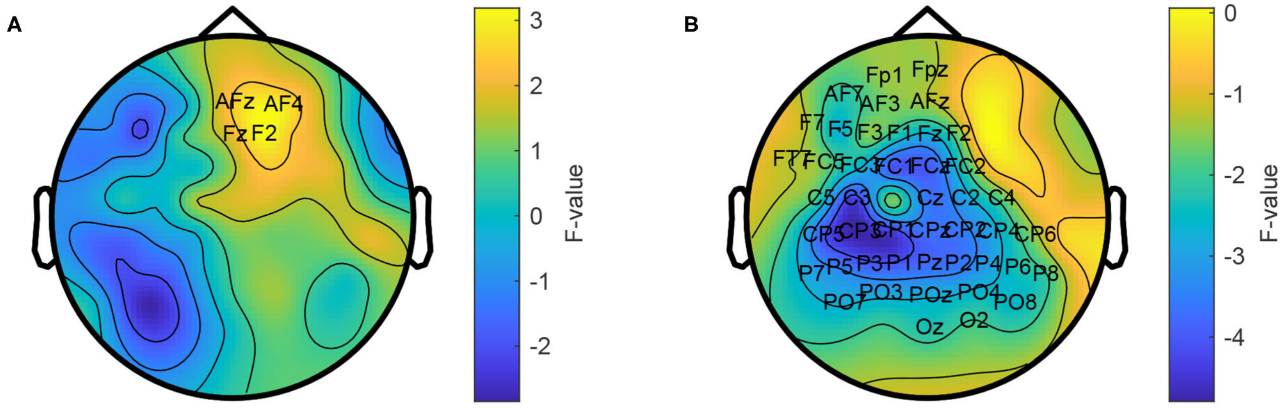


FIGURE 3 | Topographic comparison of power spectra in the collaborative and individual condition. Power in the frequency range from 66 to 120 Hz is stronger in the collaborative condition **(A)** and weaker in the range from 1 to 30 Hz **(B)**. The average power across the respective frequency ranges is shown. Labeled electrodes indicate statistically significant differences.

at a small group of fronto-right-central electrodes ($p = 0.015$, see **Figure 3**). The maximum difference was observed at electrode F2 at 113 Hz ($F = 4.0$).

In addition, the individual condition showed a power increase in the range from 1 to 30 Hz at a left-temporo-central region ($p = 1e\text{-}3$). The maximum difference was observed at electrode Pz at 14 Hz ($F = 8.79$).

To better understand the origin of these differences, we calculated an inverse solution in the brain's 3D source space. According to this solution, the power increase in the high gamma band during the collaborative condition was located in the right superior frontal gyrus (**Figure 4A**). The power decrease in the 1–30 Hz band emerged from the right cingulate gyrus (**Figure 4C**). The source reconstruction revealed another cluster with reduced

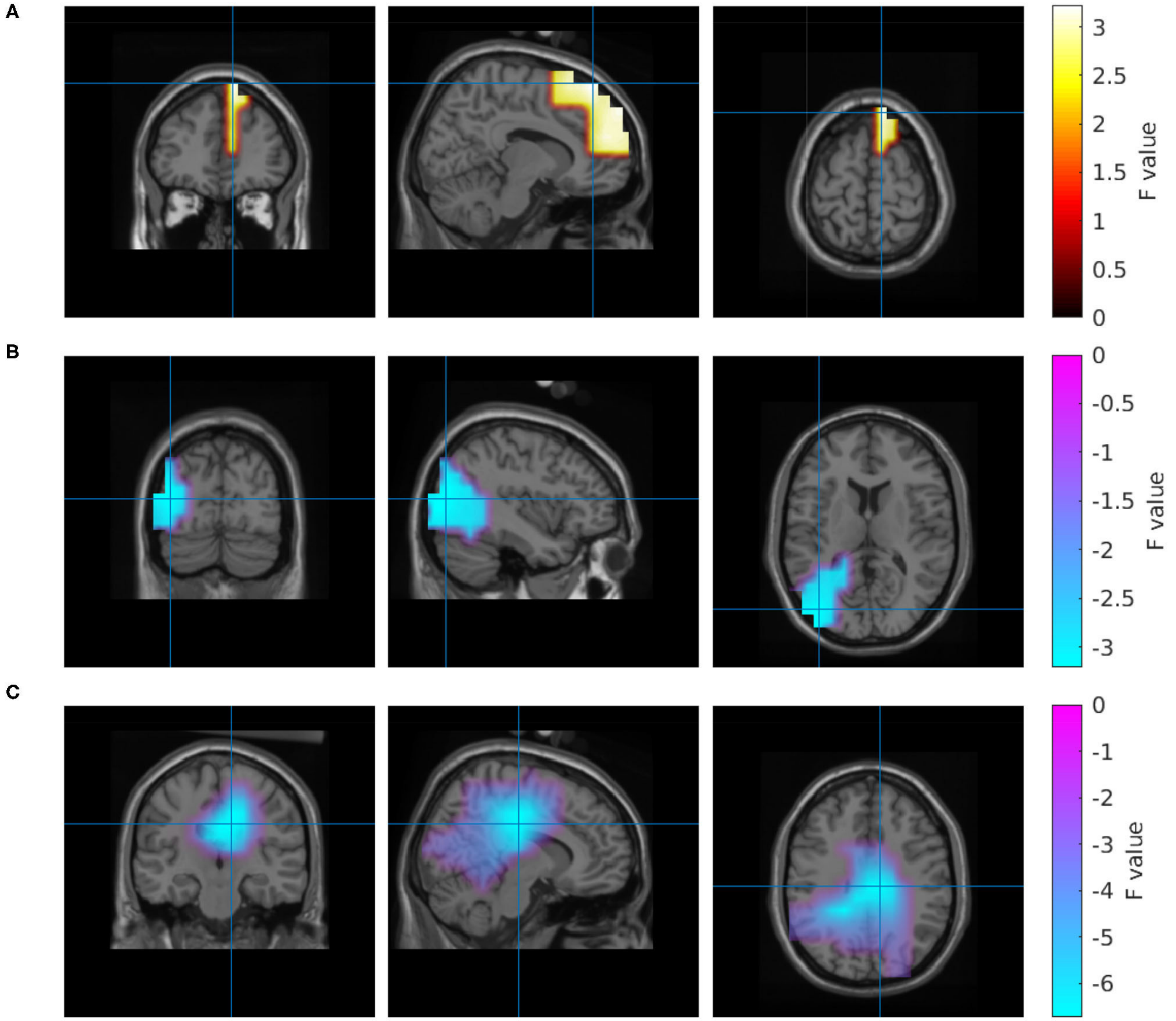


FIGURE 4 | Source reconstruction of the power differences between the collaborative and the individual condition in the frequency range 66–120 Hz **(A,B)** and 1–30 Hz **(C)**.

activity in the 66–120 Hz band in the collaborative condition which was located in the left lateral occipital cortex (**Figure 4B**). A projection of this decrease can be seen at left parietal electrodes around P5 in the topography (**Figure 3A**); however, the statistics of this condition difference was above threshold there. **Table 1** summarizes the frequencies, locations, and statistics of the power difference sources.

3.3. Hyper-Brain Coupling

In order to search for signs of coordinated brain activity in the dyads, we analyzed the EEG data from the partners using the following coupling methods: amplitude coupling (AC), power envelope coupling (PC), partial directed coherence (PDC), directed transfer function (DTF), Granger causality (GC), coherence (COH), and phase-locking value (PLV). None of the methods indicated a systematic increase or decrease of coupling between the two conditions. The *p*-values of the respective cluster-based randomization statistics are listed in **Table 2**.

3.4. Correlating Behavioral Data and Source Activity

We finally analyzed possible relations between the individual brain activity and the respective behavioral parameters of the participant. To this end, we reconstructed the activity in 3D source space and transformed it to the frequency domain. We then correlated the power spectrum at each voxel with the tracking error along the axis of the participant, motion energy and the three ratings R1–3.

TABLE 1 | Contrasting brain activity in the collaborative and individual conditions.

Freq. (Hz)	Region of maximal correlation	MNI			*p*-value	F statistic
66–120	Right superior frontal gyrus A9l, lateral area 9	10	40	60	0.033	3.23
66–120	Left lateral occipital cortex mOccG, middle occipital gyrus	−40	−80	10	0.033	−3.21
1–30	Right cingulate gyrus A23v, ventral area 23	10	−20	40	0.001	−6.76

TABLE 2 | Contrasting the collaborative and individual conditions using different coupling methods.

Connectivity measure	Collab>indiv	Indiv>collab
AC	0.13	0.51
PC	0.3	0.87
PDC	0.6	0.54
DTF	0.71	0.73
GC	0.07	–
COH	0.3067	0.2947
PLV	0.1948	0.3776

p-values result from a cluster-based randomization test.

We found a negative correlation between the tracking error and beta-band power in a region in the left inferior parietal lobule. Properties of this correlation are listed in **Table 3**, and a visualization of this region is shown in **Figure 5**.

For the subjective experience of own performance (R1), the analysis revealed a more complex pattern of regions with correlated brain activity. A small frontal region, a larger occipital region and two left and right temporal regions comprise the set of brain regions that were positively correlated with ratings of the own performance. Whereas the strongest correlation in the occipital regions was observed in the alpha band, the remaining regions had their maximum correlation in the beta band. In addition, a negative correlation was detected in the right hemisphere of the cerebellum in the delta frequency range. Properties of these clusters are listed in **Table 4**, and **Figure 6** visualizes their location and extension.

The activity in four regions correlated with the ratings of collaboration (R3). Alpha-band activity in a right parieto-occipital region showed the maximum correlation. Similar to ratings of own performance, the experience of collaboration also correlated with activity in left and right temporal regions. In contrast to all other clusters, this correlation was not frequency-specific and could be observed in the range from 20 to 120 Hz. Again, a negative correlation at delta frequencies was found in the right hemisphere of the cerebellum. A quantitative description of the correlations with R3 is given in **Table 5**, and the regions are visualized in **Figure 7**.

There were no correlations with motion energy ($p > 0.14$) or ratings of the partner's performance performance ($p > 0.1$) found.

4. DISCUSSION

4.1. Joint Action or Joint Attention?

The joint target tracking task required the two players to coordinate their actions in space and time and therefore complies with a working definition of joint action (Sebanz et al., 2006). It has to be pointed out though that the non-redundant control of the ball along orthogonal axes completely decoupled both agents' action and effect spaces. A feature which in our view is crucial in joint action, the adjustment of actions to those of the other agents, may therefore be missing in this paradigm. Hence interpretations of the results on the background of joint action should be taken with a grain of salt. Moving the virtual ball together with the partner in the collaborative condition and controlling it by oneself in the individual condition clearly should have switched between joint attention in the former and individual attention in the latter. Contrasting both conditions therefore can shed light on the neural processes involved in joint attention as a preliminary stage in joint action (Maye et al., 2017).

4.2. (No) Hyper-Brain Coupling in Arrhythmic Interaction

We applied a set of coupling analysis methods which have been used in the literature to reveal temporal coordination in hyper-activity data. None of them detected changes in hyper-brain coupling when partners switched between solving the task on their own and solving it together. One explanation for this apparently disappointing finding may be that the manipulation of the social context was just not strong enough to detect hyper-brain coupling. We think it is difficult to explain then, however, why switching between joint and individual target tracking should entail weaker changes in social coupling than, for example, switching strategies (cooperation/competition, collaborating/defecting) in card or economic games, e.g., (Babiloni et al., 2007b; De Vico Fallani et al., 2010), or in sports games (Liu et al., 2021). An alternative explanation could be that our paradigm did not impose rhythms which could be modulated by social context. It seems that some researchers also considered this possibility in their studies. For example, in Lindenberger et al. (2009) the authors discuss, that "... given that the reported rhythms were all in the low EEG frequency range, one plausible explanation could be that the similarities in sensorimotor feedback (at least partially) contributed to the inter-brain synchronization" (Konvalinka and Roepstorff, 2012). Thus, if it would not have been the similarity of the sensorimotor feedback as such but its rhythmic components[3], then a lack of rhythmicity in the sensory feedback may explain our difficulties to observe modulations of the hyper-brain synchronization in the EEG in our paradigm. A recent study of hyper-brain activity in a virtual tennis game may round up this conclusion. Amplitudes of alpha- and beta-band oscillations were correlated when players

---

[3] The study (Lindenberger et al., 2009) investigated guitarists playing a short melody together.

TABLE 3 | Clusters of brain activity correlating with tracking error.

Freq.	Region of maximal correlation	MNI	Corr. coeff.	*p*-value	F statistic
18 Hz	Left inferior parietal lobule A40rd, rostrodorsal area 40 (PFt)	−40 −50 50	−0.21	0.001	−5.9151

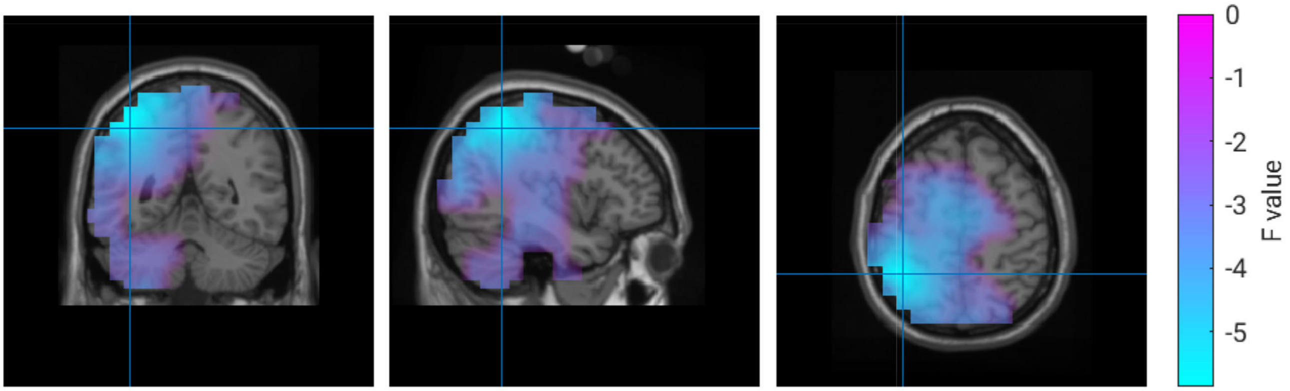


FIGURE 5 | Correlation between 18 Hz power and tracking error.

TABLE 4 | Clusters of brain activity correlating with rating of own performance.

Freq. (Hz)	Region of maximal correlation		MNI		Corr. coeff.	*p*-value	F statistic
10	Right precuneusA7m, medial area 7 (PEp)	0	−80	40	0.18	0.007	4.99
25	Left precentral gyrus A4ul, area 4 (upper limb region)	−40	−20	70	0.17	0.023	4.64
22	Right inferior parietal lobule A40rd, rostrodorsal area 40 (PFt)	60	−20	30	0.16	0.042	4.48
22	Right superior frontal gyrus A9m, medial area 9	0	40	40	0.16	0.042	4.48
1	Cerebellum	40	−50	−30	−0.17	0.011	−4.86

TABLE 5 | Clusters of brain activity correlating with rating of collaboration.

Freq. (Hz)	Region of maximal correlation		MNI		Corr. coeff.	*p*-value	F statistic
9	Right inferior parietal lobule A39rv, rostroventral area 39	40	−70	740	0.23	0.02	4.66
20–120	Left precentral gyrus A4hf, area 4 (head and face region)	−60	10	30	0.22	0.036	4.49
1	Cerebellum	40	−40	−30	−0.22	0.058	−4.45

were in the same team playing doubles (cooperative condition), but they were anti-correlated when playing in opposing teams (competitive condition) (Liu et al., 2021). Although the authors acknowledge the possibility that the observed coupling could be a by-product of the interaction in the game, they argue that the manipulation of the social context should prevail. Hence, a closer analysis of the effect of the interaction dynamics on the results in hyper-brain studies in general seems advisable.

4.3. Individual-Brain Signatures of Collaboration

More support for a successful manipulation of social context comes from the observation that tracking the target together or individually very well induced changes of the neuronal activities in the individual brains. When participants collaborated, we found increased high-gamma-band activity of a region in the right superior gyrus, which has been linked to explicit emotional processing (Iaria et al., 2008). At the same time, these oscillations were reduced in the lateral occipital cortex. Gamma oscillations in this region have been attributed to visual object processing, and they were modulated by attention and expectation (Tallon-Baudry et al., 2004). In addition, oscillations in the beta and lower frequency bands were also reduced under the collaborative condition in the cingulate gyrus, which likewise has been shown to be involved in visuomotor integration (Field et al., 2015) and social emotions (Britton et al., 2006). The topography of this decrease is similar to the result from a previous study on joint attention (Lachat et al., 2012), where the maximal modulation appeared between 11 and 13 Hz. We conclude that the two modes of solving the tracking task induced differences in visuomotor integration processes, emotional processing, and attention.

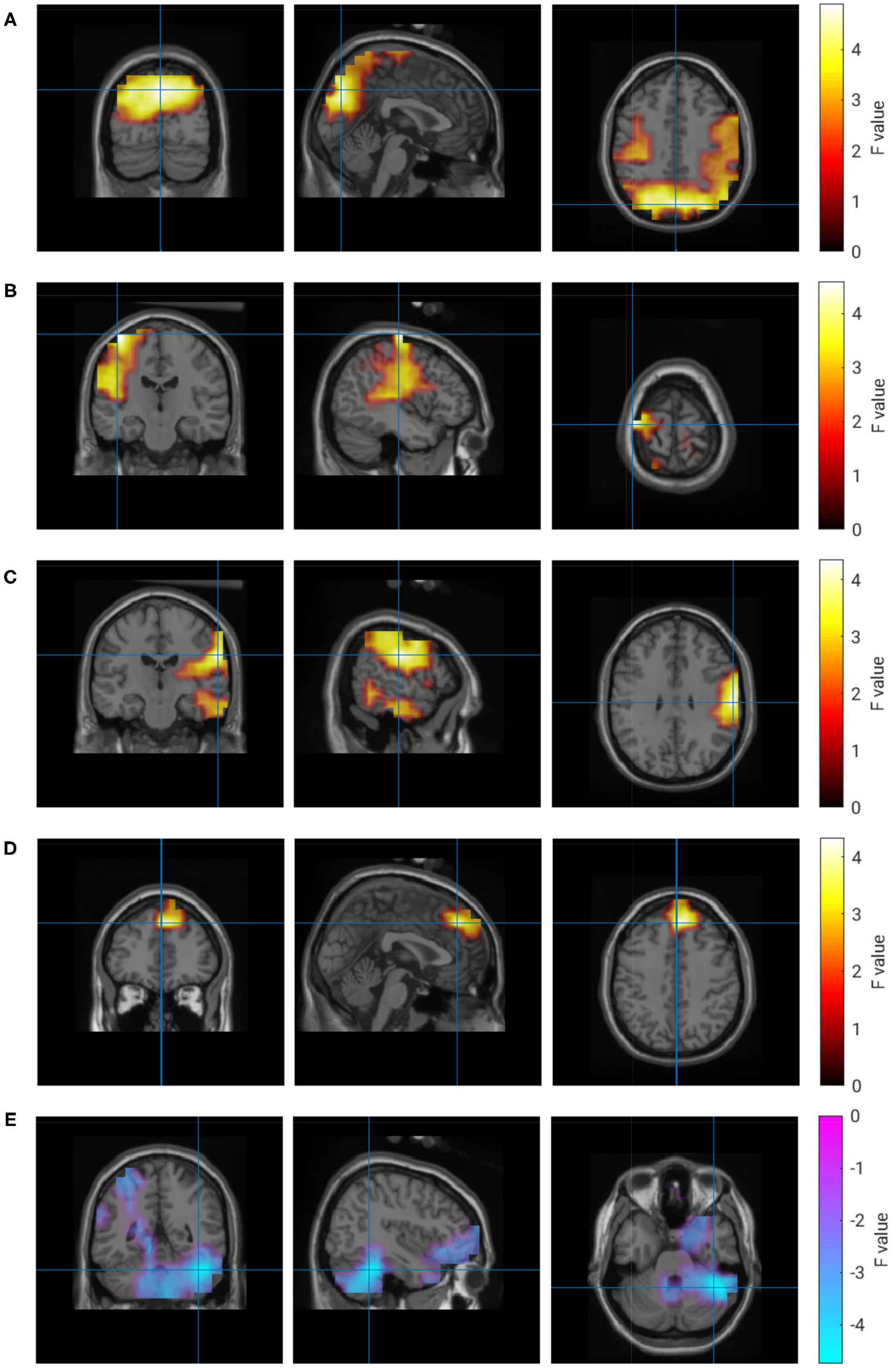


FIGURE 6 | Correlation of brain activity and rating of own performance. **(A)** 10 Hz, **(B)** 25 Hz, **(C)** 22 Hz, **(D)** 22 Hz, **(E)** 1 Hz.

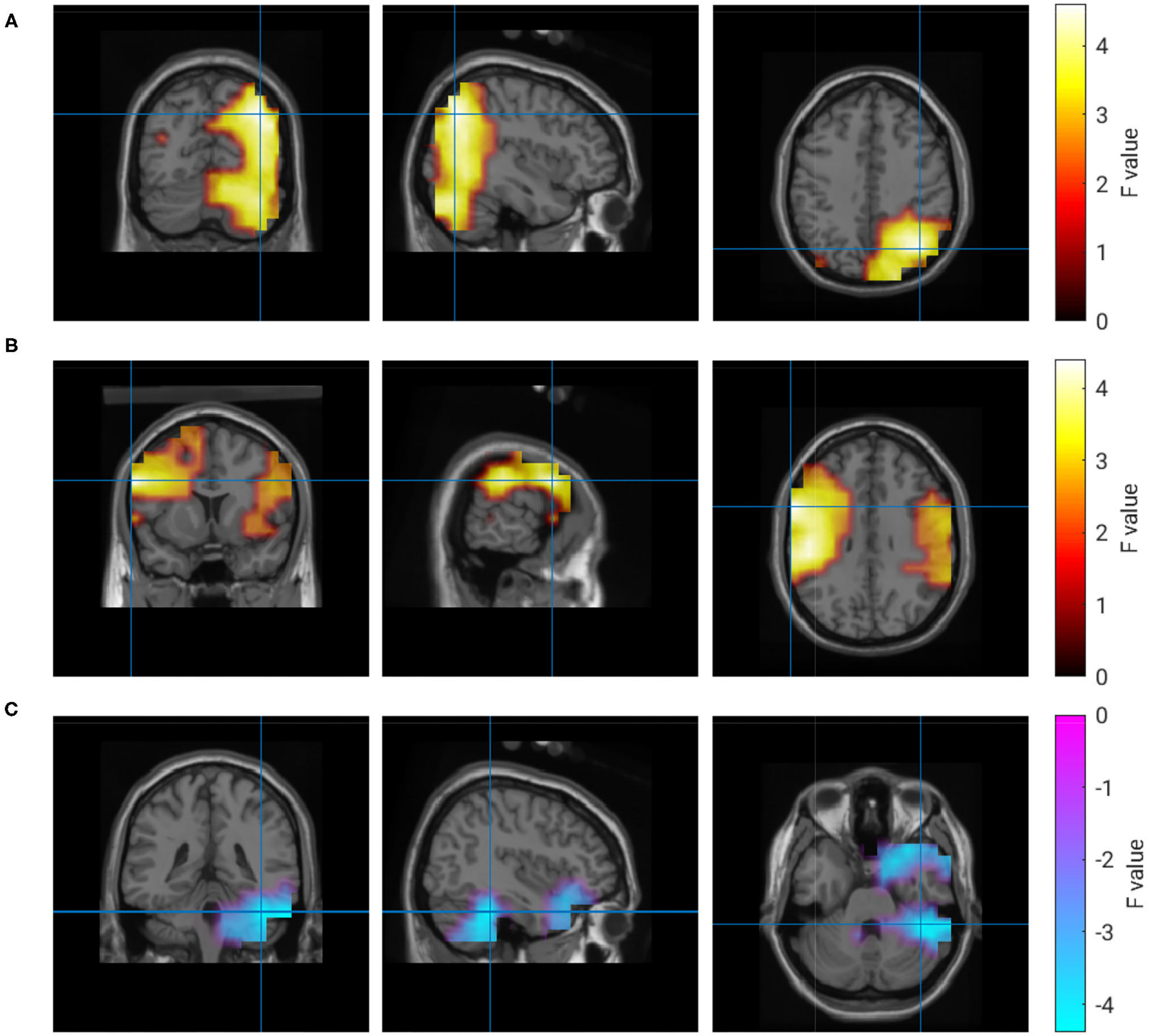


FIGURE 7 | Correlation of brain activity and rating of the partner's performance. **(A)** 9 Hz, **(B)** 20–120 Hz, **(C)** 1 Hz.

4.4. Neural Correlates of Task Performance and Self-Assessment

We also found several patterns of brain activity which were related to task performance and subjective experience. The tracking error showed a negative correlation with activity in the left inferior parietal lobule, i.e., stronger activity in this area was associated with better task performance. Activity in this region has been found for target motion prediction (Kawawaki et al., 2006), action execution, observation and imagination (Lacourse et al., 2005; Molinari et al., 2012; Jiang et al., 2015), linking it to the human mirror system (Dinstein et al., 2007). The correlation with task performance was specifically with activity in the beta band. This frequency band has been traditionally regarded as an idling rhythm in the motor system (Pfurtscheller et al., 1996), but newer accounts confer it a more active role for the maintenance of steady-state force output and a more efficient processing of proprioceptive feedback needed for monitoring the status quo and recalibrating the sensorimotor system (Engel and Fries, 2010). It may therefore well exhibit the "active akinetic process" that controls the miniature movements of the index finger to keep the ball on the target. Since beta rhythms are also related to the expectation of upcoming events, possibly they may also have been induced by the players waiting for the target to reverse its movement direction.

Whereas our analysis approach yielded a single activity cluster which correlated with the objective task performance, it revealed more complex spatio-spectral structures for the self-assessment of task performance. Taking into account the weak agreement between behavioral indicators of objective and subjective performance, i.e., tracking error and rating of own performance, this distinctiveness may come as no surprise. The maximum correlation occurred in the precuneus, a brain area which interestingly has shown activity for reflective self-awareness (Kjaer et al., 2002) and representation of the mental self (Lou et al., 2004; Cavanna and Trimble, 2006). But the cluster of significant correlation with ratings of task performance extends over large parts of the superior parietal lobule, which is considered a node in the default mode network. Interestingly, activity specifically in the alpha band, like in our study, showed significant overlap with the default mode network for self-referential thoughts and during a social game task (Knyazev et al., 2011). The researchers hypothesized that synchronization of internal mental processes as opposed to the processing of external stimuli might be the primary function of alpha oscillations in this region.

We found three additional clusters in which beta-band activity correlated with the outcome of the self-evaluation. One of these extended over the left motor cortex and therefore is likely related to movements of the right index finger. Activity in the inferior parietal lobule has been linked to motor representations of finger movements (Gerardin et al., 2003) and, as part of the human mirror neuron system, action observation and execution (Arnstein et al., 2011). The frontal activation cluster seems to match with the anteromedial portion of the right superior frontal gyrus, a region which again is part of the default mode network as well as the cognitive control network (Li et al., 2013). It may be interesting to note that the same region showed a stronger activation in the 66–120 Hz band during collaboration, whereas the correlation with self-assessment was found only for activity around 22 Hz. This may result from different neuronal populations with different activity profiles located in the same region or from the same population exhibiting a functional segregation by different frequency bands. In any case this comes as a reminder that the spatial and spectral activation profiles should be seen in an integrated fashion.

Whereas all clusters discussed so far were positively correlated with the self-assessment of performance, a region with negative correlation was located in the cerebellum. Traditionally, the cerebellum has been considered a site where models of the motor apparatus reside, and which are used for predicting the consequences of actions (Wolpert et al., 1998). But the cerebellum is also engaged in the acquisition and discrimination of sensory information (Gao et al., 1996), sensorimotor coordination, prediction and error-based learning, and affective socio-cognitive processing (Sokolov et al., 2017). With respect to tracking the target by finger movements in our paradigm, we think the cerebellar cluster can be closely linked with other studies which found that executed as well as imagined hand movements cause activity in the cerebellum (Lacourse et al., 2005), and that motor activity of and sensory signals from the fingers are mapped in the cerebellum (Wiestler et al., 2011). This link is further supported by the finding that the delta-band EEG has information which can be used to decode finger movements (Paek et al., 2014). Recently cerebellar activity has been linked to social cognition (Van Overwalle et al., 2020), and the correlation with the subjective performance evaluation could result from the observed goal-directed body movements of the partner in the context of our paradigm.

4.5. Neural Correlates of Intersubjectivity

Like for own performance, the experience of the success of collaborating with the partner was also correlated with activity in the left precentral gyrus. The extension of this cluster in the right hemisphere is very similar to the cluster around the right IPL that correlated with ratings of own performance. Whereas the spatial distribution of these clusters is similar between the two ratings, they differ in their frequency specificity. For ratings of collaboration, correlations can be observed in the beta band as well as across all of the gamma range, but for ratings of own performance, the correlation is specifically in the beta band.

The involvement of cerebellar activity also resembles the correlation structure for performance ratings. Recalling that ratings of own performance and collaboration were highly correlated, it would be interesting to know whether this was the result of the similarities in the neuronal activation profiles or whether our method yielded similar results because the input, i.e., the ratings, were correlated. Unfortunately it is not possible to answer this question with the approach selected for this study.

Despite these similarities, the spatial distribution of the alpha-band cluster exhibits notable differences between the two ratings. For ratings of collaboration, correlations were found only in the posterior part of the parietal lobe of the right hemisphere, extending into the ipsilateral occipital lobe, whereas it extends over both hemispheres for ratings of own performance.

The absence of correlations with motion energy and ratings of the partner's performance suggests two conclusions: First, the correlations with the other two ratings are not simply the result of how much the participants moved the tablet, at least not to a significant extent. And second, the evaluation of the partner's performance did not systematically covary with activity in the brain regions for motor control, sensorimotor integration and emotional processing of the own body like it did for evaluating the own performance or collaboration. This is insofar surprising as all three ratings were significantly correlated within the individuals. It is unlikely that the ratings were randomly given either, because they showed a stronger correlation within the dyad than the other two. We therefore conjecture a form of coupling of the partners in mutually rating their performance which our analysis methods were not able to pick up.

4.6. Action and Subjective Experience May Share the Same Neuronal Processes

Seen from a bird's eye perspective our analyses revealed several clusters in brain regions that are known to be involved in motor control, sensory processing, sensorimotor integration, and executive control. If one accepts that the observed correlations indicate, at least in parts, a causal relation, then the conclusion is that self-assessment of performance and collaboration are significantly modulated by the neuronal processes that govern sensorimotor coordination during the target-tracking task. This interpretation is supported by the finding that in the majority of clusters, oscillatory activity specifically in the alpha and beta band correlates with subjective experience. From the range of functional significance assigned to the two frequency bands, there is one aspect that sees alpha and beta oscillations together controlling task performance: motor inhibition. Following the target not only requires tilting the tablet in the right direction, but also involves suppressing unwanted movements and a great deal of precision in the motor control, both of which is not possible without inhibition. Stronger alpha oscillations may indicate better inhibitory control and tighter timing of cortical processing (Klimesch et al., 2007). Likewise, beta oscillations may be related to the maintenance of the sensorimotor set and the suppression of unexpected external events (Engel and Fries, 2010). It may also be hypothesized that the clustered activity in the alpha, beta and high gamma band are an index for cognitive operations of the global neuronal workspace (Palva and Palva, 2007); however, this would require showing phase coordination of these oscillations.

It has to be pointed out that the discussed brain regions have shown activity in many other tasks and contexts, and we selected the studies we deemed the most related to the experimental paradigm we investigated here. Nevertheless, the activation patters seem to match well with the cognitive requirements for solving the task. Together with the finding that the clusters were associated with specific frequency bands, the alternative interpretation of the observed correlations as sheer covariation seems less likely. Our analyses therefore support the view that the subjective experience of social interaction involves the interaction of distributed neuronal populations, many of which are considered controlling motor execution and coordinating sensorimotor processing. What's more, physiological processes in the body as indexed by autonomic parameters like heart rate variability, skin conductance and breathing rhythm also have been shown to be informative about experience of performance and collaboration (Maye et al., 2020). An integrated analysis of activity in the cerebral and autonomic nervous system, though extremely complex, may be a necessary next step toward a deeper understanding of the body for the emergence of intersubjectivity.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the Medical Association of the City of Hamburg. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

The experiments were conceived and designed by AE, Mircea Stoica (see section Acknowledgments), and AM. The experiments were performed by Mircea Stoica and AM. The data were analyzed by AM and TW. The paper was written by AM, TW, and AE. All authors contributed to the article and approved the submitted version.

FUNDING

The work described in this paper was supported by the EU through project 'socSMCs' (H2020 – 641321) and by the DFG through project TRR 169/B1.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to Mircea Stoica for supporting the development of the setup and data recording.

REFERENCES

Arnstein, D., Cui, F., Keysers, C., Maurits, N. M., and Gazzola, V. (2011). μ-suppression during action observation and execution correlates with BOLD in dorsal premotor, inferior parietal, and SI cortices. *J. Neurosci.* 31, 14243–14249. doi: 10.1523/JNEUROSCI.0963-11.2011

Babiloni, F., Astolfi, L., Cincotti, F., Mattia, D., Tocci, A., Tarantino, A., et al. (2007a). "Cortical activity and connectivity of human brain during the prisoner's dilemma: an EEG hyperscanning study," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Lyon: IEEE), 4953–4956. doi: 10.1109/IEMBS.2007.4353452

Babiloni, F., Cincotti, F., Mattia, D., Fallani, F. D. V., Tocci, A., Bianchi, L., et al. (2007b). "High resolution EEG hyperscanning during a card game," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Lyon: IEEE), 4957–4960. doi: 10.1109/IEMBS.2007.4353453

Baccalá, L. A., and Sameshima, K. (2001). Partial directed coherence: a new concept in neural structure determination. *Biol. Cybernet.* 84, 463–474. doi: 10.1007/PL00007990

Baccalá, L. A., Sameshima, K., and Takahashi, D. (2007). "Generalized partial directed coherence," in *2007 15th International Conference on Digital Signal Processing* (Wales: IEEE), 163–166. doi: 10.1109/ICDSP.2007.4288544

Britton, J. C., Phan, K. L., Taylor, S. F., Welsh, R. C., Berridge, K. C., and Liberzon, I. (2006). Neural correlates of social and nonsocial emotions: an fMRI study. *Neuroimage* 31, 397–409. doi: 10.1016/j.neuroimage.2005.11.027

Burgess, A. (2013). On the interpretation of synchronization in EEG hyperscanning studies: a cautionary note. *Front. Hum. Neurosci.* 7:881. doi: 10.3389/fnhum.2013.00881

Cavanna, A. E., and Trimble, M. R. (2006). The precuneus: a review of its functional anatomy and behavioural correlates. *Brain* 129, 564–583. doi: 10.1093/brain/awl004

De Jaegher, H., Di Paolo, E., and Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends Cogn. Sci.* 14, 441–447. doi: 10.1016/j.tics.2010.06.009

De Vico Fallani, F., Nicosia, V., Sinatra, R., Astolfi, L., Cincotti, F., Mattia, D., et al. (2010). Defecting or not defecting: how to "read" human behavior during cooperative games by EEG measurements. *PLoS ONE* 5:e14187. doi: 10.1371/journal.pone.0014187

Dinstein, I., Hasson, U., Rubin, N., and HEEGer, D. J. (2007). Brain areas selective for both observed and executed movements. *J. Neurophysiol.* 98, 1415–1427. doi: 10.1152/jn.00238.2007

Dodel, S., Cohn, J., Mersmann, J., Luu, P., Forsythe, C., and Jirsa, V. (2011). "Brain signatures of team performance," in *International Conference on Foundations of Augmented Cognition* (Orlando, FL: Springer), 288–297. doi: 10.1007/978-3-642-21852-1_35

Dodel, S., Pillai, A., Fink, P., Muth, E., Stripling, R., Schmorrow, D., et al. (2010). "Observer-independent dynamical measures of team coordination and performance," in *Motor Control: Theories, Experiments, and Applications,* *1st Edn*., eds F. Danion and M. Latash (Oxford University Press), 72–101. doi: 10.1093/acprof:oso/9780195395273.003.0004

Engel, A. K., and Fries, P. (2010). Beta-band oscillations–signalling the status quo? *Curr. Opin. Neurobiol*. 20, 156–165. doi: 10.1016/j.conb.2010.02.015

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., et al. (2016). The human brainnetome atlas: a new brain atlas based on connectional architecture. *Cereb. Cortex* 26, 3508–3526. doi: 10.1093/cercor/bhw157

Field, D. T., Inman, L. A., and Li, L. (2015). Visual processing of optic flow and motor control in the human posterior cingulate sulcus. *Cortex* 71, 377–389. doi: 10.1016/j.cortex.2015.07.014

Fuchs, T., and Jaegher, H. (2009). Enactive intersubjectivity: participatory sense-making and mutual incorporation. *Phenomenol. Cogn. Sci*. 8, 465–486. doi: 10.1007/s11097-009-9136-4

Gao, J.-H., Parsons, L. M., Bower, J. M., Xiong, J., Li, J., and Fox, P. T. (1996). Cerebellum implicated in sensory acquisition and discrimination rather than motor control. *Science* 272, 545–547. doi: 10.1126/science.272.5261.545

Gerardin, E., Lehericy, S., Pochon, J.-B., Montcel, S. T., d., Mangin, J.-F., et al. (2003). Foot, hand, face and eye representation in the human striatum. *Cereb. Cortex* 13, 162–169. doi: 10.1093/cercor/13.2.162

Hamilton, A. (2020). Hyperscanning: beyond the hype. *Neuron* 109, 404–407. doi: 10.1016/j.neuron.2020.11.008

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science* 303, 1634–1640. doi: 10.1126/science.1089506

Iaria, G., Committeri, G., Pastorelli, C., Pizzamiglio, L., Watkins, K. E., and Carota, A. (2008). Neural activity of the anterior insula in emotional processing depends on the individuals' emotional susceptibility. *Hum. Brain Mapp*. 29, 363–373. doi: 10.1002/hbm.20393

Jiang, D., Edwards, M. G., Mullins, P., and Callow, N. (2015). The neural substrates for the different modalities of movement imagery. *Brain Cogn*. 97, 22–31. doi: 10.1016/j.bandc.2015.04.005

Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., Mckeown, M. J., Iragui, V., et al. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* 37, 163–178. doi: 10.1111/1469-8986.3720163

Kaminski, M. J., and Blinowska, K. J. (1991). A new method of the description of the information flow in the brain structures. *Biol. Cybernet*. 65, 203–210. doi: 10.1007/BF00198091

Kawawaki, D., Shibata, T., Goda, N., Doya, K., and Kawato, M. (2006). Anterior and superior lateral occipito-temporal cortex responsible for target motion prediction during overt and covert visual pursuit. *Neurosci. Res*. 54, 112–123. doi: 10.1016/j.neures.2005.10.015

Kjaer, T. W., Nowak, M., and Lou, H. C. (2002). Reflective self-awareness and conscious states: PET evidence for a common midline parietofrontal core. *Neuroimage* 17, 1080–1086. doi: 10.1006/nimg.2002.1230

Klimesch, W., Sauseng, P., and Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Res. Rev*. 53, 63–88. doi: 10.1016/j.brainresrev.2006.06.003

Knyazev, G. G., Slobodskoj-Plusnin, J. Y., Bocharov, A. V., and Pylkova, L. V. (2011). The default mode network and EEG alpha oscillations: an independent component analysis. *Brain Res*. 1402, 67–79. doi: 10.1016/j.brainres.2011.05.052

Konvalinka, I., and Roepstorff, A. (2012). The two-brain approach: how can mutually interacting brains teach us something about social interaction? *Front. Hum. Neurosci*. 6:215. doi: 10.3389/fnhum.2012.00215

Lachat, F., George, N., et al. (2012). Oscillatory brain correlates of live joint attention: a dual-EEG study. *Front. Hum. Neurosci*. 6:156. doi: 10.3389/fnhum.2012.00156

Lachaux, J.-P., Rodriguez, E., Martinerie, J., and Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Hum. Brain Mapp*. 8, 194–208.

Lacourse, M. G., Orr, E. L., Cramer, S. C., and Cohen, M. J. (2005). Brain activation during execution and motor imagery of novel and skilled sequential hand movements. *Neuroimage* 27, 505–519. doi: 10.1016/j.neuroimage.2005.04.025

Li, W., Qin, W., Liu, H., Fan, L., Wang, J., Jiang, T., et al. (2013). Subregions of the human superior frontal gyrus and their connections. *Neuroimage* 78, 46–58. doi: 10.1016/j.neuroimage.2013.04.011

Lindenberger, U., Li, S.-C., Gruber, W., and Müller, V. (2009). Brains swinging in concert: Cortical phase synchronization while playing guitar. *BMC Neurosci*. 10:22. doi: 10.1186/1471-2202-10-22

Liu, H., Zhao, C., Wang, F., and Zhang, D. (2021). Inter-brain amplitude correlation differentiates cooperation from competition in a motion-sensing sports game. *Soc. Cogn. Affect. Neurosci*. 16, 552–564. doi: 10.1093/scan/nsab031

Lou, H. C., Luber, B., Crupain, M., Keenan, J. P., Nowak, M., Kjaer, T. W., et al. (2004). Parietal cortex and representation of the mental self. *Proc. Natl. Acad. Sci. U.S.A*. 101, 6827–6832. doi: 10.1073/pnas.0400049101

Lübbert, A., Göschl, F., Krause, H., Schneider, T. R., Maye, A., and Engel, A. K. (2021). Socializing sensorimotor contingencies. *Front. Hum. Neurosci*. 15:542. doi: 10.3389/fnhum.2021.624610

Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024

Maris, E., Schoffelen, J.-M., and Fries, P. (2007). Nonparametric statistical testing of coherence differences. *J. Neurosci. Methods* 163, 161–175. doi: 10.1016/j.jneumeth.2007.02.011

Maye, A., and Engel, A. K. (2016). *Extending Sensorimotor Contingencies to Cognition*. Strüngmann Forum Reports. MIT Press.

Maye, A., Isern-Mas, C., Barone, P., and Michael, J. A. (2017). Sensorimotor accounts of joint attention. *Scholarpedia* 12:42361. doi: 10.4249/scholarpedia.42361

Maye, A., Lorenz, J., Stoica, M., and Engel, A. K. (2020). Subjective evaluation of performance in a collaborative task is better predicted from autonomic response than from true achievements. *Front. Hum. Neurosci*. 14:234. doi: 10.3389/fnhum.2020.00234

Molinari, E., Baraldi, P., Campanella, M., Duzzi, D., Nocetti, L., Pagnoni, G., et al. (2012). Human parietofrontal networks related to action observation detected at rest. *Cereb. Cortex* 23, 178–186. doi: 10.1093/cercor/bhr393

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). Fieldtrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci*. 2011:156869. doi: 10.1155/2011/156869

Oostenveld, R., Stegeman, D. F., Praamstra, P., and van Oosterom, A. (2003). Brain symmetry and topographic analysis of lateralized event-related potentials. *Clin. Neurophysiol*. 114, 1194–1202. doi: 10.1016/S1388-2457(03)00059-2

Paek, A. Y., Agashe, H., and Contreras-Vidal, J. L. (2014). Decoding repetitive finger movements with brain activity acquired via non-invasive electroencephalography. *Front. Neuroeng*. 7:3. doi: 10.3389/fneng.2014.00003

Palva, S., and Palva, J. M. (2007). New vistas for α-frequency band oscillations. *Trends Neurosci*. 30, 150–158. doi: 10.1016/j.tins.2007.02.001

Pascual-Marqui, R. D. (2007). Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: exact, zero error localization. *arXiv preprint arXiv:0710.3341*.

Pascual-Marqui, R. D., Michel, C. M., and Lehmann, D. (1994). Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *Int. J. Psychophysiol*. 18, 49–49. doi: 10.1016/0167-8760(84)90014-X

Pfurtscheller, G., Stancák, A., and Neuper, C. (1996). Post-movement beta synchronization: a correlate of an idling motor area? *Electroencephalogr. Clin. Neurophysiol*. 98, 281–293. doi: 10.1016/0013-4694(95)00258-8

Sänger, J., Müller, V., and Lindenberger, U. (2012). Intra-and interbrain synchronization and network properties when playing guitar in duets. *Front. Hum. Neurosci*. 6:312. doi: 10.3389/fnhum.2012.00312

Sebanz, N., Bekkering, H., and Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends Cogn. Sci*. 10, 70–76. doi: 10.1016/j.tics.2005.12.009

Sokolov, A. A., Miall, R. C., and Ivry, R. B. (2017). The cerebellum: adaptive prediction for movement and cognition. *Trends Cogn. Sci*. 21, 313–332. doi: 10.1016/j.tics.2017.02.005

Tallon-Baudry, C., Bertrand, O., Hénaff, M.-A., Isnard, J., and Fischer, C. (2004). Attention modulates gamma-band oscillations differently in the human lateral occipital cortex and fusiform gyrus. *Cereb. Cortex* 15, 654–662. doi: 10.1093/cercor/bhh167

Tognoli, E., Kovacs, A., Suutari, B., Afergan, D., Coyne, J., Gibson, G., et al. (2011a). Behavioral and brain dynamics of team coordination part I: task design. *Lect Notes Comput. Sci*. 1, 257–264. doi: 10.1007/978-3-642-21852-1_32

Tognoli, E., Kovacs, A., Suutari, B., Afergan, D., Coyne, J., Gibson, G., et al. (2011b). Behavioral and brain dynamics of team coordination part II: neurobehavioral performance. *Lect. Notes Comput. Sci.* 1, 376–382. doi: 10.1007/978-3-642-21852-1_44

Van Overwalle, F., Manto, M., Cattaneo, Z., Clausi, S., Ferrari, C., Gabrieli, J. D., et al. (2020). Consensus paper: cerebellum and social cognition. *Cerebellum* 19, 833–868. doi: 10.1007/s12311-020-01155-1

Vesper, C., and Sebanz, N. (2016). "Acting together: representations and coordination processes," in *Shared Representations: Sensorimotor Foundations of Social Life*, eds S. S. Obhi and E. S. Cross (Cambridge: Cambridge University Press), 216–235. doi: 10.1017/CBO9781107279353.012

Wiestler, T., McGonigle, D. J., and Diedrichsen, J. (2011). Integration of sensory and motor representations of single fingers in the human cerebellum. *J. Neurophysiol.* 105, 3042–3053. doi: 10.1152/jn.00106.2011

Wolpert, D. M., Miall, R., and Kawato, M. (1998). Internal models in the cerebellum. *Trends Cogn. Sci.* 2, 338–347. doi: 10.1016/S1364-6613(98)01221-2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read
for greatest visibility
and readership

FAST PUBLICATION
Around 90 days
from submission
to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative,
and constructive
peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers
acknowledged by name
on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data
and methods to enhance
research reproducibility

DIGITAL PUBLISHING
Articles designed
for optimal readership
across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics
track visibility across
digital media

EXTENSIVE PROMOTION
Marketing
and promotion
of impactful research

LOOP RESEARCH NETWORK
Our network
increases your
article's readership