Towards a Model of the Interplay of Mentalizing and Mirroring in Embodied Communication

Sebastian Kahl (skahl@uni-bielefeld.de)
Social Cognitive Systems Group, Faculty of Technology
Bielefeld University, Inspiration 1, 33619 Bielefeld, Germany

Stefan Kopp (skopp@uni-bielefeld.de)
Social Cognitive Systems Group, Faculty of Technology
Bielefeld University, Inspiration 1, 33619 Bielefeld, Germany

Abstract

The social brain contains distinct networks for mentalizing as well as for mirroring-based action observation. We present work towards a model of how these two systems may interact during embodied communication. The model connects a mentalizing system for attributing and inferring different orders of belief about own and other's mental states, with a hierarchical predictive model for online action perception and production. We discuss interactions between both systems and describe simulation experiments in which two agents equipped with this model engage in embodied communication. Results demonstrate how mentalizing affords higher robustness of communication by enabling interactive grounding processes.

Keywords: Cognitive Modeling; Social Cognition; Mirroring; Mentalizing; Coordination; Gesture; Embodied Virtual Agents

Introduction

A growing body of research has started to study the cognitive mechanisms of social interaction and communication in humans. Two partially overlapping networks have been identified in the "social brain" (Van Overwalle, 2009): an *action observation system* for perceiving and recognizing others' behavioral cues, and a *mentalizing system* for understanding others in terms of attributed mental states or theory of mind (ToM). Both systems have been studied a lot separately. Action observation is nowadays widely assumed to rest upon principles of prediction-based processing (A. Clark, 2013). This means that predictions about expected sensory stimuli (either caused by an observed stimuli or through self-generated action) are continuously formed and evaluated against incoming sensory input to determine a prediction error that informs further processing. A core mechanism to derive such predictions are covert simulations of the observed behavior, based on own sensorimotor action representations that are assumed to be shared between perception processes and action production processes. This is what is also referred to as *mirroring* in behavior perception. The principle of prediction-based processing has also been argued to account for language production and comprehension (Pickering & Garrod, 2013) or the social brain more generally (Frith & Frith, 2010).

Yet, what is less clear is the full picture of how the mentalizing system and the mirroring system work together: How does behavior perception change when a behavior is assumed to be "for me", i.e., intended to be socially meaningful? How are perception-action couplings modulated by social interaction and mentalizing? How do these processes enable coordination devices in communication like feedback, joint attention, or grounding?

While a number of mechanisms have been hypothesized, these questions are far from being answered. What is undisputed is that interacting with other agents assumed to be "intentional" is fundamentally different from interacting with non-intentional things or objects (Gangopadhyay & Schilbach, 2012). That is, intentionality-attribution and underlying mentalizing influence sensory processing to become "social perception" (Wykowska, Wiese, Prosser, & Müller, 2014; Teufel, Fletcher, & Davis, 2010). A key component to trigger this intentional stance is social attention, most prominently signaled through gaze (Teufel et al., 2010). For example, Ciaramidaro, Becchio, Colle, Bara, and Walter (2014) recently found that social gaze leads to the attribution of communicative intent, which in turn differentially recruits the mirror neuron system and mentalizing system networks in processing the behavior of the (now considered) interlocutor. Clearly, these processes play an important role not only in the solitary observation events in which they have been studied mostly so far, but even more so in continuous online interaction (Myllyneva & Hietanen, 2015; Schilbach et al., 2013).

In this paper, we present work towards a model of how a mentalizing system interacts with a mirroring-based action observation system in continuous online interaction. In the remainder of this paper, we first review related modeling attempts and then present an integrated model that formalizes the two systems in terms of computational processes, as well as their roles and dynamic interplay in inter-agent communication. We report results from simulations of embodied communication with hand gesture, gaze, and head nods/shakes, between two virtual agents each of which equipped with the integrated model. We analyze how different abilities for mentalizing enable increasingly complex social coordination, from mere mimicry to eventually shared understanding.

Related work

So far there have only been few attempts to combine mentalizing, perception and action control in dynamic social interaction. In their MOSAIC model, Wolpert, Doya, and Kawato (2003) underline that a true communicative model needs to close the communicative loop and must be perceptive to the

observer's responses and ultimately her understanding. In their model, a hierarchy of paired forward and inverse models is hypothesized as a basic mechanism for processing movement as well as beliefs or intentions. Sadeghipour and Kopp (2011) present the Empirical Bayesian Belief Update model (EBBU), a probabilistic model that implements a mirroring-based account of the perception and production of iconic gestures. In this model, a hierarchically organized representation of motor knowledge is used during action perception by forward models that formulate probabilistic expectations about possible continuations of the observed gesture. The same representation is used for action generation, with probabilistic interactions between both processes to model e.g. priming and resonance effects, and it is expanded by way of inverse models when an unknown action is encountered.

A Bayesian approach to modeling intention inference is presented by Diaconescu et al. (2014). They apply a hierarchical Gaussian filter approach to learn the intentionality of others through updating the beliefs about others' intentions during an interaction game. This framework captures individual social learning differences in order to predict the participants' perspective-taking abilities and trustworthiness. A similarly effective approach to modeling a mentalizing module was implemented by Devaine, Hollard, and Daunizeau (2014), who applied it to a gambling game scenario. Participants had to play against a Bayesian mentalizing model that could employ several levels of ToM belief attributions, from zero up to a third-order ToM. Results show that only participants who were framed to apply ToM could beat the Bayesian mentalizing model.

Few attempts have been made to clarify the interaction between mentalizing and mirroring. A meta-analysis of studies on mentalizing found that mirror areas are not recruited unless the task involves inferencing intentionality from action stimuli (Van Overwalle, 2009). Teufel et al. (2010) present the "Perceptual Mentalizing Model" which focuses on the influence of the mentalizing system on the mirror system via perceptual processing in the STS area. They differentiate between explicit and implicit theory of mind (what we call here mentalizing and mirroring, respectively) and associate the areas mPFC and TPJ to the former and the mirror neuron system to the later kind of ToM processing. Importantly, both kinds of ToM are assumed to be influenced by social sensory processing. Explicit ToM processes are associated with processing of the intentionality of a movement and have a strong influence on area STS, which acts as a *gating mechanism* to increase or decrease perception-action coupling in implicit ToM processing.

Wykowska et al. (2014) present a model of social attention, the "Intentional Stance Model", according to which the mentalizing system either exhibits a design stance or an intentional stance. The latter would be exhibited towards agents to whom mental states are attributed, while the former is applied to objects without intentionality attribution. Again, mentalizing is also associated to areas mPFC and TPJ, while social sensory processing is associated to area STS. The mentalizing system is assumed to influence sensory processing in a top-down fashion, but also to affect sensory gain control in attention mechanisms. They report the sensory gain manipulation for attentional reorienting mechanisms to be stronger in the intentional stance than in the design stance. A key aspect in triggering this intentional stance seems to be social gaze, which has been found to lead to the attribution of communicative intent (Ciaramidaro et al., 2014), which in turn differentially recruits the mirroring and mentalizing system networks in processing the behavior of the interlocutor.

Towards an integrated model

In this paper we present a novel model of how a predictive sensorimotor subsystem for action observation and production is coupled with a mentalizing subsystem for attributing mental states, to enable situated communication between embodied agents. To that end, and as described below, we embed the models in virtual agents and let them interact nonverbally to test how communicative coordination emerges from the interplay between the two subsystems (see Fig. 1).

We base our modeling approach on a number of assumptions: First, we define successful communication to be a process leading into shared communicative intentionality and established perceptual or conceptual common ground between the participants (Tomasello, 2008). This state is achieved in a dynamic grounding process (H. H. Clark & Brennan, 1991), in which communicating agents mutually present and coordinate their subjectively held beliefs about each other as well as the state of the interaction. Second, mentalizing plays a pivotal role in this through facilitating information integration and self-other distinction for coordinated action. It receives information from, and affects the mirroring subsystem, which itself processes perceived action in an immediate fashion and feeds into higher-level mentalizing. Third, we assume that coordinated action in communication highly depends on the order of ToM realized by the mentalizing subsystem. 2nd order reasoning, i.e. beliefs about other-beliefs, is minimally necessary for any cooperative behavior that goes beyond accidental coordination. Finally, gaze plays a special

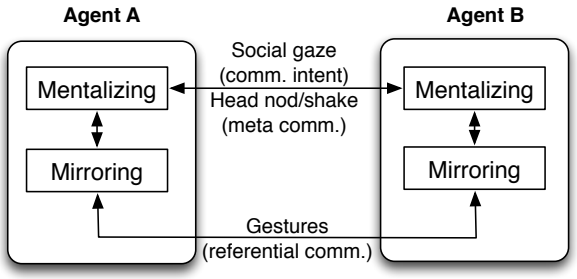

Figure 1: Outline of the overall structure of the simulated communication between two cognitive embodied agents.

role in signaling and regulating social attention and communicative intent. Mirror areas were found to be recruited especially when intentional action is expected (Van Overwalle, 2009) and one indicator for communicative intention is social gaze (Ciaramidaro et al., 2014). We hence assume that gaze triggers mentalizing and thus also mirroring activity. Staying within the confines of the nonverbal domain, we also include head-nods and head-shakes as feedback signals for agreement and disagreement.

Mentalizing subsystem

The mentalizing subsystem is a model of an agent's subjective ToM, which processes definite information about itself and needs to infer others' mental states from perceptual input. A detailed depiction of the mentalizing subsystem is given in Figure 2, which we will refer to and describe in more detail in the *Simulation results* section. In its current version, the subsystem utilizes a rather simple set of inference heuristics to model how mental state attributions arise and change in social interaction. In detail, this mentalizing model consists of three sets of mental state attributes for different orders of ToM reasoning: Beliefs held about mental states of myself (*me*) or the interlocutor (*you*) constitute what we call 1st order ToM. In pursuit of a minimal cognitive model of mentalizing, we assume that only one order of ToM higher is needed for what we want to model. In contrast to the classical recursively nested beliefs (beliefs about beliefs about...), however, we stipulate these beliefs to be held about mental states that both interlocutors have in common (*we*). This is what we call 2nd-order ToM. Generally, mental states consist of beliefs, desires, and intentions. The functional role that we ascribe to 2nd order ToM is to keep track of common ground, the desire to agree, and the collaborative state of communication more generally.

Mirroring subsystem

We adopt the Empirical Bayesian Belief Update model (EBBU) (Sadeghipour & Kopp, 2011) for action observation and production. It implements a probabilistic hierarchical representation of sensorimotor knowledge about iconic gestures, along with basic prediction, evaluation and activation processes that are used both in perception and generation of gestures. On the lowest level *motor commands* are stored that represent single movement segments in time and space. Hand trajectories are given as directed graphs with their edges representing the motor commands. On the intermediate level, *motor programs* represent paths in the motor command graph. In that way each motor program stands for a meaningful movement. The highest level of abstraction stores *motor schemas* that cluster similar motor programs to represent functional equivalence classes (e.g. depicting similar shapes). When observing a gesture trajectory, probabilistic forward models predict expectations about possible continuations of the observed gesture by running internal simulations on all levels of the hierarchy in parallel. While simulating, the hierarchical motor knowledge structure "resonates" to the observed gesture performance. In each time step, the model's predictions are compared to the actual perception in order to evaluate the corresponding motor commands, programs, or schemas. When no corresponding representation of the observed behavior can be found in the knowledge structure, inverse models can execute a motor learning mechanism that imparts the novel behavior into the motor command graph and corresponding motor program and schema, when no similar schema exists.

Our model of a sensorimotor subsystem employs this EBBU model and is equipped with knowledge about schemas, programs and commands of different gesture trajectories for 'circle', 'square', 'surface' and 'waving'. Those were learned from real human motion data. Single motor programs for a schema can take up to five seconds to produce, with motor commands being activated every tenth of a second. For every new observation of a hand trajectory entering the subsystem, posterior probability distributions are updated using the EBBU rule (Sadeghipour & Kopp, 2011). The topmost level distributions about motor schemas are taken as a proxy for a gesture's meaning, and are linked to 1st order mental state attributes in the mentalizig subsystem.

Integration and interplay

Our goal is to integrate mentalizing with mirroring-based action perception to account for how behavior and mental states arise and interact dynamically in a communicative interaction.

In other models of mentalizing (Teufel et al., 2010; Wykowska et al., 2014), the detection of social gaze plays a crucial role. In our model, it triggers an inference for communicative intent to the gazing agent, which in turn affects further processing of its behavior. In particular, as long as the intention to communicate can be inferred, for any observed gesture processed by the mirroring subsystem, the most likely motor schema hypothesis is immediately projected into the mentalizing subsystem where it forms a mental state attributed as a *you*-belief. This resembles the gating mechanism of area STS as suggested by Teufel et al. (2010). Correspondingly, a *me*-belief would cause the mirroring subsystem to recruit the intended motor schema for production. The current version of the mirroring subsystem is only capable of processing hand gestures; gaze and head movements are thus directly asserted to the mentalizing subsystem.

Depending on the degree of ToM available in the agent's mentalizing subsystem, communicative intent can trigger an inferred desire to reach mutual agreement about the understanding of the produced gesture. This is assessed by applying a threshold for *good-enough* understanding to the likelihood of beliefs about mental states of *me* and *you* (1st order ToM). Note, however, when this threshold is exceeded the producer agent still cannot be certain about the correct understanding in the recipient unless sufficient feedback is provided. Here, we require at least one correct reproduction of the gesture. Further, head-shake and head-nod signals are employed for meta-communication and can either increase or

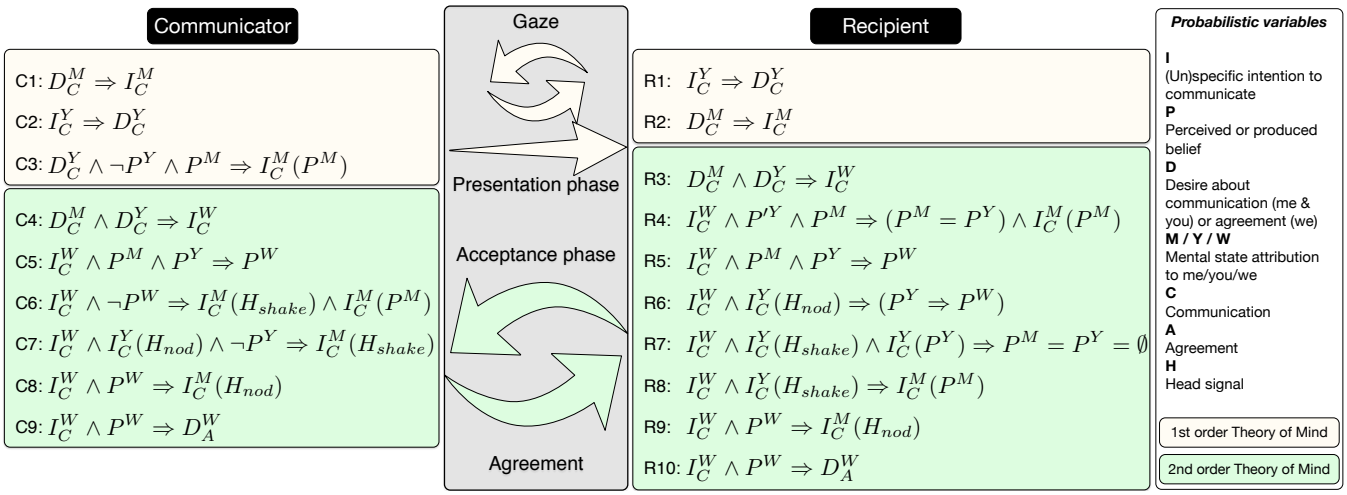

Figure 2: Attributes and inference heuristics in the mentalizing subsystem applied during different phases of the interaction. The basis for complex inference is "Communicative Intention", inferred from social gaze. The "Communicator" agent enters the "Presentation Phase", followed by an "Acceptance Phase" of interactive grounding, where higher order mental attributions are needed for both agents to reach "Agreement".

decrease confidence in the respective *you*-belief.

Simulation results

In order to test the model in online interactions we implemented the model and ran simulations with two virtual agents, each of which equipped with its own integrated model. At the start of the simulation, both agents only have a predefined set of mental states about themselves. They can communicate using four gestures ('circle', 'square', 'surface', and 'waving') that are perceived and generated as 3D hand trajectories, as well as head nods/shakes that are transferred as simple timed key-value pairs. Gestures are produced with a configured amount of white noise, normalized to the maximum movement vectors in the motor schema, so that 10% noise reflect only a small amount of deviation during gesturing. The amount of noise, the ability for 2nd order ToM, and the good-enough threshold for minimal confidence in observing a gesture are the independent variables to parametrize the simulation.

We ran six simulation setups: 10%/20%/30% noise with enabled or disabled 2nd order ToM capacity, and a static confidence threshold of 0.8. Each of the setups was run 100 times, always with identically configured agents. Simulations ended either when both agents believed to have reached agreement, or without 2nd order ToM, as soon as the *Communicator* finished its gesture production. As dependent variables we collected the probability distribution of the attributed *you*-belief about a gesture's meaning after every processing of a hand trajectory. We were particularly interested in the effects that different degrees of mentalizing have on the inter-agent coordination dynamics. The complexity of the communication depends on inferred communicative intent, signaled via social gaze. As soon as mutual communicative intent is established, the simulation follows a typical grounding process with presentation and acceptance phases (H. H. Clark & Brennan, 1991), where the *Communicator* always starts with producing a 'circle' gesture.

To examplify the effect of the mental attributions and inferences possible in 1st and 2nd order ToM, Figure 3 illustrates two typical interaction patterns from our simulation study. The overt behavior of two agents, a *Communicator* and a *Recipient*, are shown along with the inferences drawn after perceiving or producing a certain behavior, with indices referring to the inference rules as shown in Figure 2.

The interaction at the bottom shows a sequence of behavior and inferences typical for 1st order ToM mentalizing. The configured desire to communicate triggers *rule C1*, hence gaze behavior is perceived by the *Recipient* (*rule R1*). Since the *Recipient* is equally configured, its reciprocal gaze behavior (*rule R2*) triggers an inference about the *Recipient's* desire to communicate in the *Communicator* (*rule C2*), and consequently a gesture is produced (*rule C3*).

The interaction at the top shows behavior and inferences enabled through 2nd order ToM. While in the beginning there is a similarity to the 1st order ToM interaction, additionally *rules R3 and C4* are triggered and establish the agents' common communicative intent and thus the foundation for meaningful coordination behavior. After the initial gesture production the *Recipient's* mirroring subsystem provides the mentalizing subsystem with the most likely interpretation for the *Communicator's* behavior. That novel behavior triggers *rule R4*, by which the *Recipient* would ideally produce the understood gesture back to the *Communicator*, but in this interaction the gesture was understood with a likelihood above the good-enough threshold. This triggers *rule R5 and R9* as well, leading to a head-nod. Since the *Communicator* has no idea what the *Recipient* has understood the head-nod behavior is

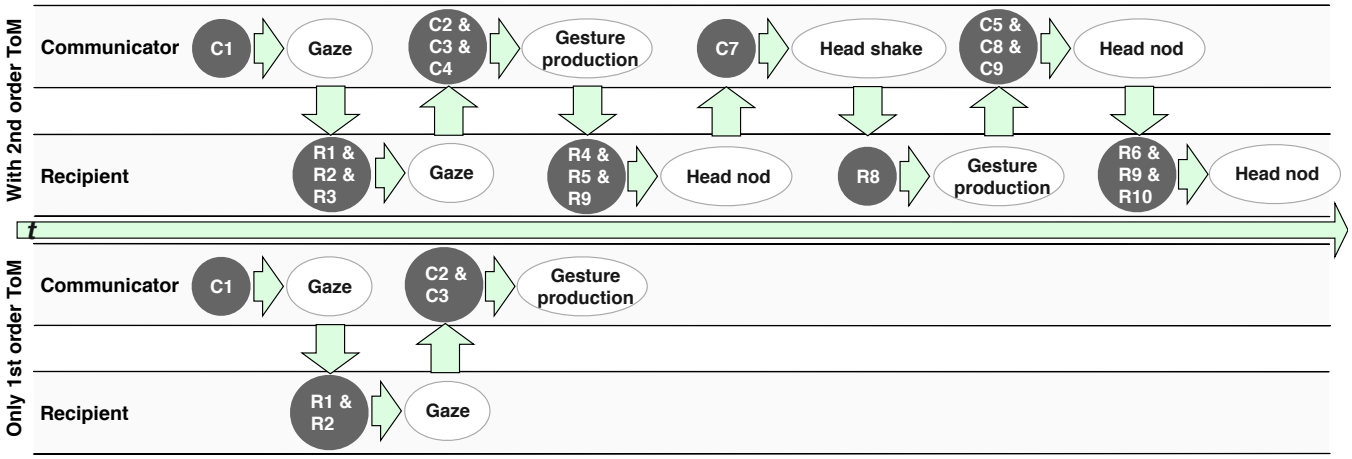

Figure 3: Example interactions from our simulation when both agents have 2nd order ToM (top) or 1st order ToM (bottom). Overt behavior is shown along with the triggered mentalizing inferences (gray circles; indices referring to Figure 2).

answered by a head-shake (*rule C7*), which triggers the *Recipient* to produce its understood gesture back to the *Communicator* (*rule R8*). The *Communicator* understands the gesture, which triggers rules equivalent to those in the *Recipient* (*rule C5, C8, and C9*), leading to a head-nod, which is equivalently answered by the *Recipient* (*rule R6, and R9*) and finalizes the interaction through mutually believed agreement (*rule R10*).

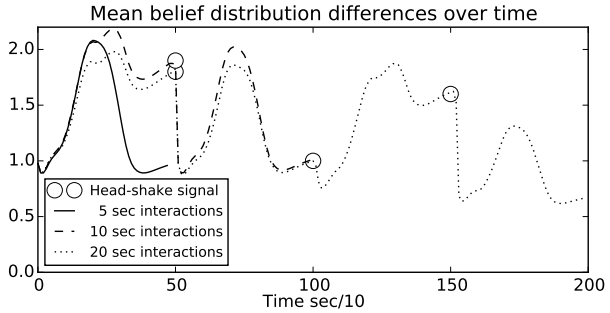

Figure 4: Simulations show KL-divergence between agents' beliefs during interactions of different extent, averaged over noise and ToM conditions.

To test the agents' ability to coordinate with and without 2nd order ToM enabled, we analyzed the Kulbach-Leibler Divergence between the probability distributions of the *Recipient's you*-belief and the *Communicator's me*-belief, i.e. the "target belief". Figure 4 shows the divergence over interaction time. Without 2nd order ToM only one gesture was produced within 5 seconds. With 2nd order ToM the duration was strongly dependend on the correct understanding of observed gestures. The more mistakes, due to noise, the more correction effort emerged and hence longer interactions. Analyzed were interactions with length of at least 10 seconds and 20 seconds, respectively. These plots show the average success of coordination, especially in longer interactions. To test

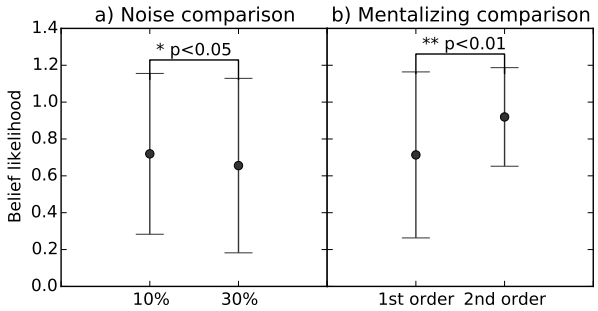

Figure 5: Analyses show a) mean differences between noise conditions after 5 sec., and b) mean differences between final likelihood about another's belief in ToM conditions.

the effect of noise we compared the success of both agents reaching the target belief within last three hand position samples before the 5 second mark, averaged over ToM conditions (Figure 5(a)). The comparison shows a significant difference ($t(1198) = 2.4$, $p < 0.05$, $d = 0.14$) between 10% ($M = 0.6$, $SD = 0.4$) and 30% ($M = 0.7$, $SD = 0.5$) noise conditions on gesture understanding. Subsequently, we tested the influence of 2nd order ToM, also by analyzing the success of reaching the target belief (Figure 5(b)). A comparison of the final beliefs averaged over all noise conditions with 2nd order ToM ($M = 0.9$, $SD = 0.27$) and without ($M = 0.7$, $SD = 0.45$) showed that 2nd order ToM leads to significantly more likely success in coordination ($t(598) = 6.8$, $p < 0.01$, $d = 0.56$).

Conclusions

In this paper we have investigated the questions how mere action observation needs to be complemented by higher order mentalizing, and how those systems need to interact in order to account for the dynamic inter-agent coordination mechanisms that are required for successful communication. Our view is based on the notion that there is a strong difference

in the interaction with intentional, versus non-intentional entities. To that end we augmented a predictive action observation system with a 'minimal' mentalizing model that enables distinct mental perspectives corresponding to beliefs about *me*, *you*, and *we*.

Our approach is to explicate possible interactions between mentalizing and mirroring in terms of computational process models that can be implemented and tested in actual simulations of dynamic unfolding interactions. Here, we investigated whether 1st order mental state attributions are sufficient to infer the information necessary to successfully act towards a communicative goal, or whether higher order theory of mind can give a distinct advantage. Our results demonstrate that mentalizing affords interactive grounding and thus makes communication significantly more robust and efficient. However, even with higher order mentalizing capacities, a too large perturbation of the communicative signals led to long interaction times due to the inefficient error correcting mechanism emerging from both agents' goal for successful communication. Still, we established in a first prototypical modeling attempt that mentalizing is crucial for meaningful coordination behavior, and success in communication could not be established without 2nd order mental state attributes. We are thus confident that the present framework presents a good basis for further investigation of social cognitive processes, which Neuroscience is currently not yet able to elucidate. For one, we aim to provide an account for how agents dynamically compensate for noise by strategically altering their signaling behavior. Another question to pursue is how self-other distinctions manifest themselves in the sensorimotor system, e.g. when observing other agents while performing an action oneself.

Acknowledgements

This research/work was supported by the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG).

References

Ciaramidaro, A., Becchio, C., Colle, L., Bara, B. G., & Walter, H. (2014, July). Do you mean me? Communicative intentions recruit the mirror and the mentalizing system. *Social Cognitive and Affective Neuroscience*, *9*(7), 909–916. doi: 10.1093/scan/nst062

Clark, A. (2013, June). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(03), 181–204. doi: 10.1017/S0140525X12000477

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In *Perspectives on socially shared cognition* (Vol. 13, pp. 127–149). Washington, DC, US: American Psychological Association. doi: 10.1037/10096-006

Devaine, M., Hollard, G., & Daunizeau, J. (2014, December). The Social Bayesian Brain: Does Mentalizing Make a Difference When We Learn? *PLoS computational biology*, *10*(12), e1003992. doi: 10.1371/journal.pcbi.1003992

Diaconescu, A. O., Mathys, C., Weber, L. a. E., Daunizeau, J., Kasper, L., Lomakina, E. I., ... Stephan, K. E. (2014, September). Inferring on the intentions of others by hierarchical bayesian learning. *PLoS computational biology*, *10*(9), e1003810. doi: 10.1371/journal.pcbi.1003810

Frith, U., & Frith, C. (2010, January). The social brain: allowing humans to boldly go where no other species has been. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *365*(1537), 165–176. doi: 10.1098/rstb.2009.0160

Gangopadhyay, N., & Schilbach, L. (2012, July). Seeing minds: A neurophilosophical investigation of the role of perception-action coupling in social perception. *Social Neuroscience*, *7*(4), 410–423. doi: 10.1080/17470919.2011.633754

Myllyneva, A., & Hietanen, J. K. (2015, January). There is more to eye contact than meets the eye. *Cognition*, *134*, 100–109. doi: 10.1016/j.cognition.2014.09.011

Pickering, M. J., & Garrod, S. (2013, August). An integrated theory of language production and comprehension. *The Behavioral and brain sciences*, *36*(4), 329–47. doi: 10.1017/S0140525X12001495

Sadeghipour, A., & Kopp, S. (2011, September). Embodied Gesture Processing: Motor-Based Integration of Perception and Action in Social Artificial Agents. *Cognitive computation*, *3*(3), 419–435. doi: 10.1007/s12559-010-9082-z

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

Teufel, C., Fletcher, P. C., & Davis, G. (2010). Seeing other minds: Attributed mental states influence perception. *Trends in Cognitive Sciences*, *14*(8), 376–382. doi: 10.1016/j.tics.2010.05.005

Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, Massachusetts: The MIT Press. doi: 10.1353/lan.0.0163

Van Overwalle, F. (2009, March). Social cognition and the brain: a meta-analysis. *Human brain mapping*, *30*(3), 829–58. doi: 10.1002/hbm.20547

Wolpert, D. M., Doya, K., & Kawato, M. (2003, March). A unifying computational framework for motor control and social interaction. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *358*(1431), 593–602. doi: 10.1098/rstb.2002.1238

Wykowska, A., Wiese, E., Prosser, A., & Müller, H. J. (2014). Beliefs about the minds of others influence how we process sensory information. *PLoS ONE*, *9*(4). doi: 10.1371/journal.pone.0094339