Self-other distinction in the motor system during social interaction: A computational model based on predictive processing

Sebastian Kahl (skahl@uni-bielefeld.de)

Stefan Kopp (skopp@uni-bielefeld.de)

Social Cognitive Systems Group, Faculty of Technology,
Bielefeld University, Inspiration 1, 33619 Bielefeld, Germany

Abstract

During interaction with others, we perceive and produce social actions in close temporal distance or even simultaneously. It has been argued that the motor system is involved in both processes, but how does it distinguish in this processing between self and other? In this paper, we present a model of self-other distinction within a hierarchical sensorimotor system that is based on principles of perception-action coupling and predictive processing. For this we draw on mechanisms assumed for the integration of cues to generate sense of agency, i.e., the sense that an action is self-generated. We report results from simulations of different social scenarios, showing that the model is able to solve the problem of the dual use of the sensorimotor system.

Keywords: perception-action coupling; social cognition; mirroring; dual-use; sense of agency; predictive processing

Introduction

In everyday social interaction we constantly try to deduce and predict the underlying intentions behind others' social actions, like facial expressions, speech, gestures, or body posture. This is no easy problem and the underlying cognitive mechanisms and neural processes even have been dubbed the „dark matter" of social neuroscience (Przyrembel, Smallwood, Pauen, & Singer, 2012). Action recognition is commonly believed to rest upon principles of prediction-based processing (Clark, 2013), where predictions about expected sensory stimuli are continuously formed and evaluated against incoming sensory input to inform further processing. Such a predictive processing does not only inform our perception of actions of others, but also our action production in which we constantly predict the sensory consequences of our own actions and correct them in case of deviations.

Both of these processes draw on the human motor system constituting a perception-action coupling (Prinz, 1997). However, in dynamic social interaction, perception and production often need to be at work simultaneously and for both, actions of self and other. How does the sensorimotor system distinguish between self and other? And how does it interplay with higher-level cognitive processes like mentalizing to solve this social differentiation problem?

As of yet, it is not clear how exactly self-other distinction is implemented within the motor system, but there is evidence for a differentiated involvement supporting the motor system's key role in social cognition (Schütz-Bosbach, Mancini, Aglioti, & Haggard, 2006). We aim to contribute a computational modeling perspective. In previous work we devised a model of the interplay of mentalizing and prediction-based mirroring during social interaction. It demonstrated how mentalizing – even with minimal abilities to account for beliefs, desires and intentions – affords interactive grounding and makes communication more robust and efficient (Kahl & Kopp, 2015). In that work two virtual agents interacted in a communication game, each of which equipped with models of a mirroring system and mentalizing system, respectively.

In this paper we present an extension of the prediction-based model of the sensorimotor system to enable it to differentiate actions of its own from the interaction partner's actions. We start with briefly introducing the hierarchical, prediction-based model of a sensorimotor system. Then we discuss how this model can be extended to deal with concurrent perception and production in social situations. This includes a basic ability to integrate predictive and postdictive cues to form a sense of agency (SoA) that helps to differentiate between self and other. Finally, we present and discuss results from simulation studies of different simple scenarios, which test the model's ability to infer SoA for its own actions.

Computational model of a sensorimotor system

Like other attempts to model the motor system, we chose to make use of a hierarchical representation of increasing abstractions over motor commands (Wolpert, Doya, & Kawato, 2003; Sadeghipour & Kopp, 2010). In a three-level hierarchy (see Figure 1), we represent motor primitives on the lowest level (*MPrim*), followed by a motor sequence layer (*MSeq*), and motor schemas on the topmost level of abstraction (*MSchema*). Motor primitives represent single movement segments in space, motor sequences store lists of motor primitives, while motor schemas represent abstract clusters of motor sequences grouped by similarity. We assume that these representations are the basis for a prediction-based model of sensorimotor processing which underlies both action perception and production. To this end, we assume the representations to be multimodal, i.e., combining visual, motor and proprioceptive aspects of action, if available. Consequently, they are used as more or less high-level or visuomotor representations of action and their outcomes. During action perception, we further assume that the correspondence problem is solved in the sense that an observed action by another agent is mapped into one's own self-centered frame of reference. That is, we feed the perceived action trajectory directly and bottom-up into the sensorimotor system.

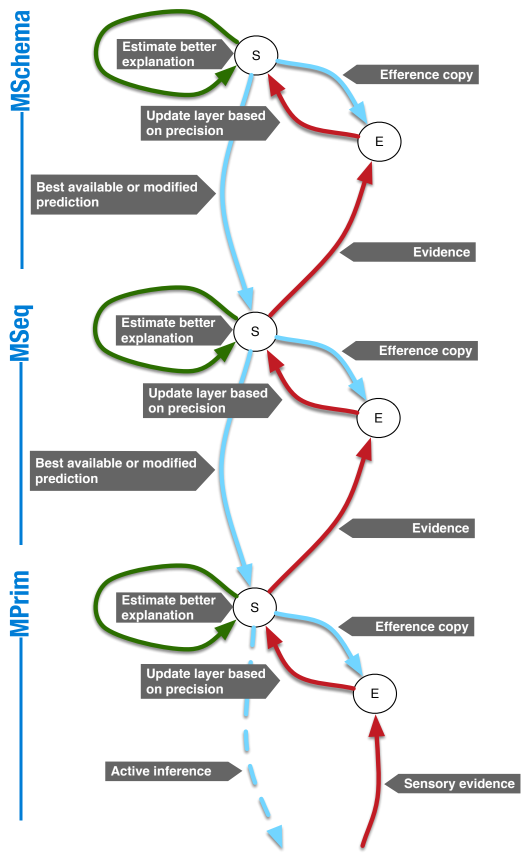

Figure 1: The predictive sensorimotor hierarchy, based on predictive processing and perception-action coupling. Predictions are sent top-down from state nodes (S) and will be compared to sensory evidence in error nodes (E) in order to drive updates within the hierarchy.

Sequence matching

In the current model motor primitives are matched against sensory input, which is assumed to consist of a sequence of the last two perceived input coordinates. Motor sequences are matched against a temporary motor sequence concurrently collected from the motor primitive layer, yielding a best match and a prediction of the next motor primitive in the sequence. Motor schemas are likewise matched against the currently best matching motor sequence. In the case of the motor primitives, before the best match is searched, the input sequence is linearly interpolated to match the length of the motor primitives and it is scaled and translated to match the motor primitive's position and size in its coordinate system. Sub-sequence matching is solved by applying euclidean distance measures, which provides high accuracy in our domain size. The same matching algorithm is used for comparing motor sequences in the motor schema layer.

Predictive sensorimotor hierarchy

The model realizes a predictive processing account resting upon assumptions of the predictive brain hypotheses (Clark, 2013). To that end, it stores representations in the form of discrete probability distributions that can be influenced both bottom-up, in the form of evidence for its last prediction from the next lower layer, and top-down in the form of a prediction by the next higher layer. Following the assumption that the main flow of information is top-down and that motor control is also just top-down sensory prediction, described as "action-oriented predictive processing", or "active inference" (Clark, 2013), all layers receive the next higher layer's prediction and evaluate it for their own bottom-up prediction in the next time step.

As shown in Figure 1, in any time step, the top layer is the first to update its discrete probability distribution in the state node (S), given its prior distribution (S_{t-1}) and the likelihood, calculated in the error node (E) based on the evidence from the layer below. The updated state node (S_t) will be used as a prediction for the current time step, influencing the layer below as a prior, and a copy will be stored in the error node for comparison in the next time step:

$$S(MSchema)_t = S(MSchema)_{t-1}E(MSeq)_{t-1}.$$

Next, the state node at the layer of motor sequences will be updated given its prior distribution, the prediction from the motor schema layer and the likelihood, calculated from the evidence in the layer below:

$$S(MSeq)_t = S(MSchema)_t S(MSeq)_{t-1}E(MPrim)_{t-1}.$$

The resulting posterior distribution will be sent as prediction to the layer below, and as evidence to the layer above. Finally, the state node at the motor primitive layer will receive an update given its prior from the last time step, the posterior from the motor sequence layer and the likelihood of the received sensory evidence (o) given all motor primitives:

$$S(MPrim)_t = S(MSeq)_t S(MPrim)_{t-1}E(o).$$

For a better understanding of the process of how the model matches the input to its hierarchical representation, see Figure 2. We have recorded handwritten capital letters using a graphical tablet. All sequences of drawing the 26 characters of the alphabet are stored with a sampling rate of 25 frames per second. From this dataset (12 primitives, mapping onto 26 sequences, mapping onto 26 schemas) we can trigger the model to draw a character, and simulate the model perceiving somebody's drawing of a character in real-time (by simply feeding the trajectories into the system as input one coordinate after the other). Figure 2 depicts three steps in the prediction-based recognition process that leads to a high probability of perceiving the drawn character.

Precision

The sensorimotor hierarchy learns motor sequences and motor schemas online, with each layer having to decide whether to add a new representation or not. One cognitively plausible

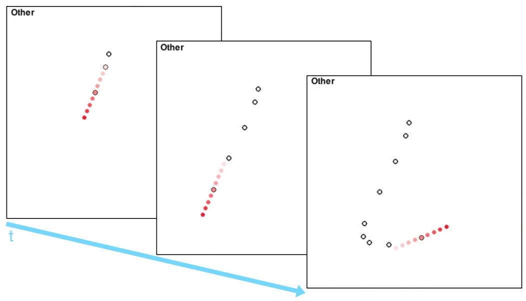

Figure 2: Prediction-based recognition process of the character L being drawn: black dots indicate input coordinates, red dotted lines represent matched motor primitives.

way to determine the need to extend upon the available motor representations is to calculate the precision of each layer's prediction against the evidence in the next time step. Friston and Frith (2015) used precision (the inverse of the variance between prior and posterior probability distributions) as a sort of cortical gain control or neuro-modulation, as to control the influence of the feedback that their Songbird models received upon their song production. We use it as a measure of how well our model can predict its environment and update or extend it according to how similar the prior ($P = S_{t-1}$) and posterior ($Q = S_t$) are:

$$Pr(P,Q) = \frac{1}{Var(Q-P)}$$

This enables each layer to evaluate its predictive power and, by thresholding, to decide whether to extend its repertoire.

Active inference

Following the assumption that overt action is basically action-oriented predictive processing (Friston, Daunizeau, Kilner, & Kiebel, 2010), we allow high-precision prediction at the layer of motor primitives to be acted out. This leads to overt and automatic *enactment* of correctly predicted actions, similar to the automatic imitation seen in patients suffering from echopraxia (Ganos, Ogrzal, Schnitzler, & Münchau, 2012). To control this motor execution, we introduced a signal into the top-layer of the hierarchy, which acts as a motor intention for a specific motor schema, including a strong boost of this motor schema's probability. This percolates down the hierarchy to boost associated representations and informs about the intention to act. Once it reaches the lowest layer of the hierarchy, and combined with the high precision threshold, the act to produce the motor representation will be allowed.

With the hierarchical model in place, we set out to find mechanisms to distinguish activations that stem from own action from those arising due to the interaction partner's action. One way is to make sure that the perceived action outcomes are correctly attributed. That is, we need to look at creating SoA, i.e., the sense that an action is self-generated.

Self-other distinction and sense of agency

How does the human brain distinguish between information about ourselves and others? Or to be more specific, how can we distinguish ourselves from others so that we do not falsely attribute an action outcome to ourselves? These questions pertain to the more general mechanisms that give rise to a sense of "feeling of control", agency, and "self". Generally, a person's SoA is believed to be influenced through predictive and postdictive (inferential) processes, which when disturbed can lead to misattributions of actions as in disorders as for example in patients suffering from schizophrenia (van der Weiden, Prikken, & van Haren, 2015). We aim to identify mechanisms in order to model these processes and their integration into a combined SoA.

The predictive process makes use of people's ability to anticipate the sensory consequences of their own actions. It allows to suppress, i.e., decrease the intensity of incoming signals which enables people to distinguish between self-caused actions and their outcomes and those actions and outcomes caused by others. One account to model these processes is based on inverse and forward models to account for disorders of awareness in the motor system and delusion of control (Frith, Blakemore, & Wolpert, 2000). This view suggests that patients can no longer link their intentions to their actions, that is they are conscious of their intention, but not of the sensory consequences of the action. As research into schizophrenia has shown, reliable and early self-other integration and distinction is important not only for the correct attribution of SoA, but also in turn for the correct attribution of intentions and emotions in social interactions. This even suggests that the attenuation of perceived sensory outcomes correlates with activation in the mirror neuron system (van der Weiden et al., 2015). Weiss, Herwig, and Schütz-Bosbach (2011) showed that there is a social aspect to predictive processes that influence SoA by comparing perceived loudness of auditory action effects in an interactive action context. They found that attenuation occured also in the interactive context, comparable to the attentuation of self-generated sound in an individual context.

The postdictive process relies more on inferences drawn after the movement in order to infer whether our intentions are contingent and consistent with the observed events (Wegner & Wheatley, 1999) and is also influenced by higher-level causal beliefs and thoughts. One important aspect of this inferential process relies on the temporal aspects of action-outcome integration. It was shown that increasing action outcome delay decreases feeling of control (Sidarus, Chambon, & Haggard, 2013). This is related to the "temporal binding window" (Colonius & Diederich, 2004), in which the sensory signals related to the outcome of an action are integrated. The width of the window is dependend on the predictability of the action outcome. Since we have more experience in predicting our own body, the temporal binding window is more narrow for own action outcomes than for other people's actions. Being able to make such a distinction allows people to monitor, infer and distinguish between causal relations for own and other's behavior. Another aspect informing the postdictive process relies on the valence attributed to an action outcome, where positive valence of an action outcome leads to stronger SoA (Yoshie & Haggard, 2013).

In sum, there are two processes that can inform SoA and hence can help to distinguish actions of self and other in social interaction. A predictive process works on the content of the action, e.g., the motor command and utilizes forward models as a mechanism to predict the to-be-produced motor command. A postdictive process works with higher-level causal beliefs like the intention to act and temporal binding as mechanisms to infer the consistency of the action outcome.

But how do these two processes work together to inform SoA and what if their cues are unreliable? Cue integration was first studied by Moore, Wegner, and Haggard (2009) who found that when predictive cues are absent external cues become more influential. Nahab et al. (2010) found in an imaging study that there is a leading and a lagging network that both influence SoA prior to and after an action. The leading network would check whether a predicted action outcome would be perceived, while the lagging network would make use of these cues to further process SoA leading to its conscious experience. It seems that in order to generate SoA, both systems do not necessarily have to work perfectly together, as there is evidence for a weighted integration of cues for agency based on their reliability (Moore & Fletcher, 2012). Also, if the reliability of the predictive process was reduced, the system put more weight on the postdictive inferential processes (Wolpe et al., 2014). Furthermore, the fluency of action-selection processes may also inform the self-other distinction, because the success of repeatedly predicting the next actions seem to accumulate over time to inform SoA (Chambon, Sidarus, & Haggard, 2014).

Modeling self-other distinction in social interaction

During online social interaction, the sensorimotor system potentially gets involved in simultaneous action perception and production processes. Our goal is to investigate how the prediction-based model can be extended to cope with the social differentiation problem during such dual-use situations. To this end, we integrate three cues from the predictive and postdictive processes into SoA for produced actions: In the predictive process, we have the match or mismatch of the predicted action-outcome. In the postdictive process, we have the intention to act and the delay in the action-outcome for temporal binding. However, it is not obvious how these cues are being integrated. As a first step, we test two simple ways to do so, namely, to connect them conjunctively or disjunctively. A conjunctive connection allows SoA to occur only if it is supported by both processes; in a disjunctive connection only one cue suffices to inform SoA, in a more flexible but potentially more error prone manner.

The predictive mechanism to differentiate self and other works based on the content of the predictions that are being sent down the hierarchy. As described above, the predictability of actions by itself provides a predictive cue for a feeling of control, or SoA. Thus, the model needs to quantify the success of a prediction about the outcome of an acted-out motor representation. Since we already have a layer's precision as a measure of success of its predictions, we can directly utilize it as a cue to model SoA.

As the postdictive inferential mechanism we model the temporal binding window as a Gaussian with its mean (μ) at the predicted delay, which is informed by the perceived action duration during learning. The Gaussian's standard deviation (σ) is scaled by the model's layer's predictive precision.

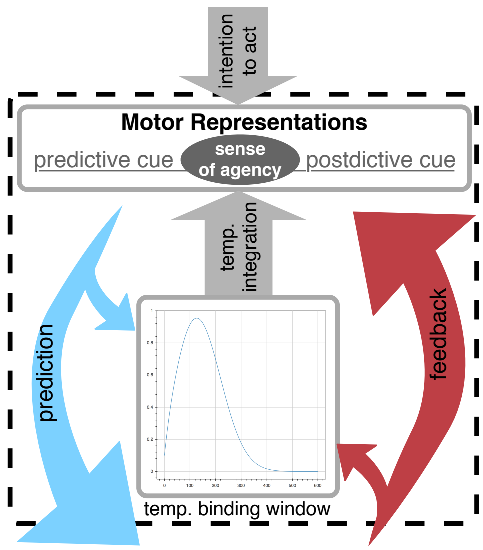

Figure 3: A model of self-other distinction based on the prediction of the consequence of an action and the postdictive integration of an intention to act with the perceived consequence of the action during a predicted temporal binding window, scaled by the predicting layer's precision.

In Figure 3 you can see a sketch of how the predictive and postdictive mechanisms work together to infer SoA for the produced action and its consequence that is perceived. The postdictive mechanism for temporal binding will be triggered by the sensorimotor hierarchy's intention to produce an action, in that it will receive a reference to the motor representation to anticipate. Information from this motor representation will then be used to model the temporal binding window. Now, when the anticipated motor representation is perceived the delay until this perception occurred (x) will be used to calculate the likelihood in the temporal binding window's Gaussian,

$$lh(\sigma,\mu,x) = \frac{1}{lh_{max}} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x-\mu^2}{2\sigma^2}},$$

with σ being scaled by precision (Pr), i.e., trust in the model's predictions. The resulting likelihood will be scaled to lh_{max}, the Gaussian's peak. This postdictive cue has to be combined with the predictive cue of the general precision of the predictions. We do this by assuming a winner takes all estimate for the predictive and postdictive cues, with a threshold at 50% probability. Postdictive and predictive cues will then be connected conjunctively **or** disjunctiveley to reach an estimate for a combined SoA. Since fluency is important for SoA (Chambon et al., 2014), we will integrate this estimate over time. The agency estimate will add to the overall SoA through a simplified Kalman filter,

$$agency_t(agency_{t-1}, agency_{estimate}) = \\ agency_{t-1} + Pr * (agency_{estimate} - agency_{t-1}).$$

By allowing the agency estimate ($agency_{estimate}$) to influence the overall SoA ($agency_t$) only through this filter, strong fluctuations are dampened and with a gain controlled by precision (Pr) the influence of the estimate will strongly depend on the success of previous predictions. This means that an agency estimate will either have a strong influence if precision is high, or SoA can only accumulate slowly if precision is low.

This is our integrated model of predictive and postdictive mechanisms which will enable the hierarchical sensorimotor system to differentiate between actions initiated from self and others.

Simulations and Results

To test the combined model's ability to solve the problem of the dual use of the sensorimotor system and differentiate between self and other we simulate a limited social situation. In this situation, the model will write a character while it either receives the correct action-outcome as feedback, or it receives delayed or different feedback than expected. This way we simulate the effect of concurrent perception of an interaction partner's action.

We test three scenarios for each combination of predictive and postdictive cues to form SoA. In the first scenario we trigger the intent to act out a motor schema and the model will perceive its own correct output as feedback. Here, the model will receive both cues correctly. In the second scenario we trigger the same intent to act, but now the model receives feedback with a delay, disrupting the postdictive cue. In the third scenario the model will again be triggered to act, while this time it receives unpredicted action-outcomes. Here, the model can receive correct cues only accidentally. The model will be triggered to produce and perceive the letter L in scenario one and two. In scenario three, the model will perceive the letter M being produced instead. We log the calculated SoA over time for each scenario.

The resulting plots in Figure 4 show SoA in the different scenarios. The upper row shows the resulting plots for the conjunctive and the lower row for the disjunctive cue integration. In scenario one, the integration of cues jumps strongly because predictive precision is high and small irregularities in timing have a strong effect. The conjunctive connection of cues does not allow for SoA in scenarios two and three, because both cues are not available simultaneously. The disjunctive connection between cues fares better, since it allows for SoA even when only one of the cues is available.

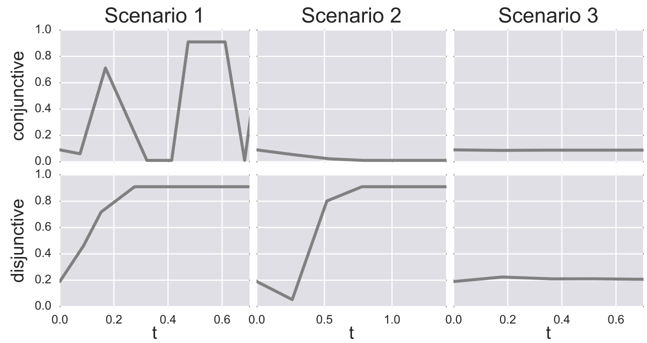

Figure 4: Resulting sense of agency over time for each scenario and each connection between predictive and postdictive cues.

In dynamic scenarios of concurrent perception and production, with feedback either from own or from other's actions, a more flexible distinction may be supported by results reported in the literature. A disjunctive integration is sound with regard to results where the reliability of the predictive process was reduced and the system put more weight on postdictive processes (Wolpe et al., 2014). Also, Moore and Fletcher (2012) found evidence for a weighted integration based on the cues' reliability. Another aspect which we found to also influence our results, was the fluent correct prediction of actions (Chambon et al., 2014). Especially in the conjunctive scenario one and the disjunctive scenario two, the accumulation of agency estimates over time was disrupted either through prediction-error or temporal binding errors.

Taken together, the results show that the model can correctly attribute its own action outcomes to itself, which contributes to distinguishing itself from an interaction partner. This differentiating role of the motor system and its strong involvement in social cognition was also proposed by (Schütz-Bosbach et al., 2006). Our cognitive model and its simulation results support this suggestion.

Conclusion

We have presented a predictive hierarchical model of the sensorimotor system, integrated with a model of self-other distinction that can solve the dual-use problem in dynamic social situations. Furthermore, we presented simulation results of different scenarios of simultaneous perception and production. We compared them to the literature on SoA and the influence of the motor system on social cognition. This comparison suggests that our model can correctly attribute SoA for its own actions, using a more flexible (disjunctive) integration of predictive and postdictive cues.

Taken together, our modeling approach supports the motor system's role in social cognition. Still, the literature on the social brain suggests that motor cognition, as well as the distinction of self and other are influenced by higher level processes, causal beliefs, and by the mentalizing network.

We agree that the interplay between mentalizing and mirroring needs to be incorporated to meet the demands of truly social systems in interaction scenarios with multiple agents. In earlier work, we already combined our previous model of the sensorimotor system with a mentalizing model in a social scenario with two virtual agents (Kahl & Kopp, 2015).

In future work, we want to improve our setup by making use of the differentiating information provided by the present model to inform higher-level cognition through an interplay with the mentalizing system. We conjecture this interplay can yield the distinction between one's own and an interaction partner's beliefs needed in social interaction, where informed reciprocity is the key to efficient and successful communication.

Acknowledgments

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

Chambon, V., Sidarus, N., & Haggard, P. (2014, May). From action intentions to action effects: how does the sense of agency come about? *Frontiers in Human Neuroscience*, *8*, 439–9.

Clark, A. (2013, June). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204.

Colonius, H., & Diederich, A. (2004, July). Multisensory Interaction in Saccadic Reaction Time: A Time-Window-of-Integration Model. *Journal of Cognitive Neuroscience*, *16*(6), 1000–1009.

Friston, K., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010, February). Action and behavior: a free-energy formulation. *Biological Cybernetics*, *102*(3), 227–260.

Friston, K., & Frith, C. (2015, November). A Duet for one. *CONSCIOUSNESS AND COGNITION*, *36*, 390–405.

Frith, C. D., Blakemore, & Wolpert, D. M. (2000, December). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *355*(1404), 1771–1788.

Ganos, C., Ogrzal, T., Schnitzler, A., & Münchau, A. (2012, July). The pathophysiology of echopraxia/echolalia: Relevance to Gilles De La Tourette syndrome. *Movement Disorders*, *27*(10), 1222–1229.

Kahl, S., & Kopp, S. (2015). Towards a Model of the Interplay of Mentalizing and Mirroring in Embodied Communication. In G. Airenti, B. G. Bara, & G. Sandini (Eds.), *Euroasianpacific joint conference on cognitive science* (pp. 300–305).

Moore, J. W., & Fletcher, P. C. (2012, March). Sense of agency in health and disease: A review of cue integration approaches. *CONSCIOUSNESS AND COGNITION*, *21*(1), 59–68.

Moore, J. W., Wegner, D. M., & Haggard, P. (2009, December). Modulating the sense of agency with external cues. *CONSCIOUSNESS AND COGNITION*, *18*(4), 1056–1064.

Nahab, F. B., Kundu, P., Gallea, C., Kakareka, J., Pursley, R., Pohida, T., ... Hallett, M. (2010, December). The Neural Processes Underlying Self-Agency. *Cerebral Cortex*, *21*(1), 48–55.

Prinz, W. (1997). Perception and action planning. *European journal of cognitive psychology*.

Przyrembel, M., Smallwood, J., Pauen, M., & Singer, T. (2012). Illuminating the dark matter of social neuroscience: Considering the problem of social interaction from philosophical, psychological, and neuroscientific perspectives. *Frontiers in Human Neuroscience*, *6*(June), 190–15.

Sadeghipour, A., & Kopp, S. (2010, November). Embodied Gesture Processing: Motor-Based Integration of Perception and Action in Social Artificial Agents. *Cognitive Computation*, *3*(3), 419–435.

Schütz-Bosbach, S., Mancini, B., Aglioti, S. M., & Haggard, P. (2006, September). Self and Other in the Human Motor System. *Current Biology*, *16*(18), 1830–1834.

Sidarus, N., Chambon, V., & Haggard, P. (2013, December). Priming of actions increases sense of control over unexpected outcomes. *CONSCIOUSNESS AND COGNITION*, *22*(4), 1403–1411.

van der Weiden, A., Prikken, M., & van Haren, N. E. M. (2015, October). Self–other integration and distinction in schizophrenia: A theoretical analysis and a review of the evidence. *Neuroscience and Biobehavioral Reviews*, *57*, 220–237.

Wegner, D. M., & Wheatley, T. (1999, July). Apparent mental causation. Sources of the experience of will. *American Psychologist*, *54*(7), 480–492.

Weiss, C., Herwig, A., & Schütz-Bosbach, S. (2011, July). The Self in Social Interactions: Sensory Attenuation of Auditory Action Effects Is Stronger in Interactions with Others. *PLoS ONE*, *6*(7), e22723–3.

Wolpe, N., Moore, J. W., Rae, C. L., Rittman, T., Altena, E., Haggard, P., & Rowe, J. B. (2014, January). The medial frontal-prefrontal network for altered awareness and control of action in corticobasal syndrome. *Brain*, *137*(1), 208–220.

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 B: Biological Sciences*, *358*(1431), 593–602.

Yoshie, M., & Haggard, P. (2013, October). Negative Emotional Outcomes Attenuate Sense of Agency over Voluntary Actions. *CURBIO*, *23*(20), 2028–2032.