

Predictive joint-action model: A hierarchical predictive approach to human cooperation

Ana Pesquita¹ · Robert L. Whitwell¹ · James T. Enns^{1,2}

Published online: 8 November 2017
© Psychonomic Society, Inc. 2017

Abstract Research in a number of related fields has recently begun to focus on the perceptual, cognitive, and motor workings of cooperative behavior. There appears to be enough coherence in these efforts to refer to the study of the mechanisms underlying human cooperative behavior as the field of joint-action (Knoblich, Butterfill, & Sebanz, 2011; Sebanz, Bekkering, & Knoblich, 2006). Yet, the development of theory in this field has not kept pace with the proliferation of research findings. We propose a hierarchical predictive framework for the study of joint-action that we call the predictive joint-action model (PJAM). The presentation of this theoretical framework is organized into three sections. In the first section, we summarize hierarchical predictive principles and discuss their application to joint-action. In the second section, we juxtapose PJAM's assumptions with empirical evidence from the current literature on joint-action. In the third section, we discuss the overall success of the hierarchical predictive approach to account for the burgeoning empirical literature on joint-action research. Finally, we consider the model's capacity to generate novel and testable hypotheses about joint-action. This is done with the larger goal of uncovering the empirical and theoretical pieces that are still missing in a comprehensive understanding of joint action.

Keywords Joint-action · Prediction · Interpersonal coordination · Social cognition

Hierarchical predictive approach to joint-action

The ability of humans to cooperate with one another vastly increases the range of their potential actions (Clark, 1996; Stix, 2014). Hence, cooperation is thought to be of central importance to our species' evolutionary success (Tomasello, 2009). In recent years, the field of cognitive science has turned its spotlight on cognition in the social milieu (Semin & Cacioppo, 2006). One consequence is that the study of the processes underlying human cooperative behavior is now a field of research in its own right, known as “joint-action” (Knoblich, Butterfill, & Sebanz, 2011). Sebanz, Bekkering, and Knoblich (2006) define joint-action as “a social interaction whereby two or more individuals coordinate their actions in space and time to bring about change in the environment.” Every joint-action, therefore, requires an interlocking of two or more individuals' intentions, actions, and perceptions (Sebanz & Knoblich, 2009). This attunement between partners is what enables ensemble musicians to create a unified sound texture and tango dancers to move together so swiftly that it seems difficult to imagine them apart.

Coordinating with others is a complex cognitive problem. Effortless as it might seem on the surface, even the simplest instances of joint action, such as playing catch or carrying an object together, require a diverse ensemble of cognitive processes to be coordinated. To successfully align ourselves with others in space and time, we need to rely on assumptions about the other person's internal processes. Current answers to this question propose that predictive processes are at the core of successful social interactions (Graziano, 2013; Manera, Schouten, Verfaillie, & Becchio, 2013; Ramnani & Miall,

✉ Ana Pesquita
anapesquita@gmail.com

¹ Department of Psychology, University of British Columbia, Douglas Kenny Building, 2136 West Mall, Vancouver, B.C. V6T 1Z4, Canada

² Department of Psychology, University of Western Australia, 33 Stirling Highway, Crawley, Western Australia 6009, Australia

2004; Sparenberg, Springer, & Prinz, 2012; Springer, Hamilton, & Cross, 2012). The idea is that we use our own cognitive resources to build mental models of other people's cognitions. Simulation of other people's cognitive states (e.g., what they are feeling, thinking, and attending) guide our expectations about their future behavior and in this way contribute to the viability of social interactions.

In the past decade, there has been much new research on the mechanisms underlying our ability to cooperate with one another. Yet, the development of theory in this field has not kept pace with the proliferation of research findings. The requirements for a theoretical architecture of joint-action have been defined by Vesper et al. (2010) as being minimally composed of representations, processes, and joint-action facilitation mechanisms. Specifically, the architecture must meet three requirements: (i) Represent a shared goal, as well as representing one's own individual contribution to the shared goal. This requirement makes explicit that the shared goal cannot be achieved through individual action. (ii) Apply monitoring and prediction processes to each partner's actions. This includes monitoring the extent to which shared goals or tasks are being fulfilled while at the same time predicting a partner's actions. (iii) Facilitate continuous coordination via coordination smoothing, defined as the process of continuously improving one's prediction of the partner's action. We note that this proposal maps the broad requirements of a joint-action model at a coarse scale; it does not specify in any detail how these requirements might be implemented.

Wolpert, Doya, and Kawato's (2003) offer a computational solution for a theory of joint-action that is premised on what they see as close parallels between individual and social motor control. These authors proposed that the sensorimotor computations involved in acting on one's own body during individual motor control are comparable to the communicative signals involved in acting on other people's behavior during social interactions. Just as individual motor control can be thought of as the process by which the central nervous system engages the musculoskeletal system to achieve a desirable end point (e.g., reaching out to pick up a cup of coffee to drink from it), in a similar way, social interaction can be thought of as motor control for the purpose of bringing about a desired change in another individual. For example, we might signal a request for more coffee by holding out our coffee cup in a forward motion toward a person who is serving coffee, thus prompting the server to fill our cup.

Wolpert et al.'s (2003) model often is cited as a useful approximation of the mechanisms sustaining joint-actions (Becchio, Sartori, & Castiello, 2010; Doerrfeld, Sebanz, & Shiffrar, 2012; Häberle, Schütz-Bosbach, Laboissière, & Prinz, 2008; Knoblich et al., 2011; Loehr, Kourtis, Vesper, Sebanz, & Knoblich, 2013; Pecenka & Keller, 2011; Ramenzoni, Sebanz, & Knoblich, 2014; Sartori, Becchio, & Castiello, 2011; Sebanz & Shiffrar, 2009; Vesper, Butterfill,

Knoblich, & Sebanz, 2010; Vesper, van der Wel, Knoblich, & Sebanz, 2013). However, in our view Wolpert et al.'s (2003) model has not garnered the full attention it deserves, perhaps because it appeared before the most recent surge of interest in joint-action. It is for the same reason—that the theory appeared slightly ahead of its time—that several key aspects of joint-action, such as goal sharing, task co-representation, and interpersonal coordination, were not addressed by Wolpert et al. (2003).

In summary, Wolpert et al.'s (2003) model does not meet the minimal requirements for a joint-action framework, as delineated by Vesper et al. (2010), just as in complimentary fashion, Vesper et al. (2010) does not match the computational rigor of Wolpert et al.'s (2003) model. In this review, we seek to bridge this gap by presenting a hierarchical predictive framework for joint-action: the predictive joint-action model (PJAM). This framework builds on Wolpert et al.'s (2003) computational notions by incorporating the necessary higher-order organization (Vesper et al., 2010) to deal with the specifics of joint-action implementation.

Fundamentals of hierarchical predictive processing

We suggest that joint-action can be best understood within a hierarchical predictive processing framework. Before presenting the proposed application of this approach to joint-action, we briefly summarize the general principles of this approach. Broadly speaking, there are two ways to implement hierarchical predictive motor control: Integrative Forward Models and Auxiliary Forward Models (Pickering & Clark, 2014). Both approaches agree that prediction, error minimization, and hierarchical modeling are core processes in motor control. Where the accounts differ is on whether these processes are implemented by dedicated mechanisms, as in Auxiliary Forward Models, or are implemented by the same mechanisms that ultimately generate the action itself, as in Integrative Forward Models.

Auxiliary forward models

The main task of motor control is to process the events needed to take an organism from its current motor state to the desired motor state. This is a complex process, because there often are several solutions that may lead to the desired outcome, the information available for selection is often distorted and delayed (due to neural-muscular limitations), and the external environment is dynamic (Wolpert, 1997).

The proposed solution involves predictive processes. In Auxiliary Forward Models, the desired motor state is converted into a motor command by generative models termed

inverse models. Inverse models receive the desired motor state and generate two outputs: a motor command that directs the musculoskeletal system during action execution, and at the same time, a copy of this motor command, termed *effference copy*, is sent to a *forward model*. The forward model simulates the musculoskeletal system and contextual environment. Taking as input the effference copy, and an estimation of the current state of the body (e.g., joint angles, movement velocities), along with the body's context (e.g., learnt dynamics of the external environment), the forward model generates predictions. It predicts how the motor system will change in response to the motor command (state predictions), and it predicts the expected sensory feedback resulting from these state changes (sensory predictions). Sensory predictions are continuously compared with sensory input signals to compute *prediction errors*. Prediction errors are used (1) to update the inverse model, which guides the selection of the next motor command, and (2) to provide a learning signal for the forward model. In this way, at each moment in time, action execution is directed towards achieving the desired motor state. Furthermore, the prediction errors are used to fine-tune the models, minimizing prediction errors (Blakemore, Goodbody, & Wolpert, 1998; Flanagan, Vetter, Johansson, & Wolpert, 2003; Wolpert, 1997; Wolpert, Ghahramani, & Jordan, 1995; Wolpert & Flanagan, 2001).

More recent conceptions of Auxiliary Forward Models, such as the hierarchical-MOSAIC account, posit that prediction and error minimization processes occur within a hierarchy of probabilistic models (Haruno, Wolpert, & Kawato, 2003). Vertically, the hierarchy is composed by different levels of action modeling, ranging from high-level symbolic representation to low-level movement feature representations. Horizontally, each level of the hierarchical-MOSAIC is composed of probabilistic models operating in parallel that generate predictions (forward models) and control signals (inverse models). To illustrate, consider the act of reaching out to grasp a coffee mug. Probabilistic models in the upper levels of the hierarchy represent information of a more abstract and symbolic nature. In this example, these models would correspond to symbolic representations of the goal of reaching for a coffee mug and the semantics that are associated with the object. Probabilistic models in the lower levels of the hierarchy represent low-level dynamics, such as movement elements, and object sensory features. In this example, these probabilistic models would represent information about limb position and velocity. Probabilistic models in the middle-levels represent different ways to structure and organize movement elements for a range of different purposes, reaching out to grasp objects being only one category. Within any given level of the hierarchy, the probabilistic models are evaluated on the basis of how well their predictions (termed *priors*) fit the signals arriving from the level

underneath (termed *responsibilities*). This bi-directional flow of information between levels permits the reentrant and recursive processes that underlie model updating and selection at different levels of the hierarchy.

Integrative forward models

For Integrative Forward Models, the desired motor state is treated as an actual state of affairs, causing a cascade of downward predictions about what should be the states of the models in the levels below. The forward models at each level in the hierarchy translate the desired motor state (e.g., higher-level representation of the action goal) into proprioceptive and exteroceptive predictions and can be thought of as representing a state probability. The predictions that are output by these probabilistic models are continuously compared against the flow of incoming information from the subordinate level in the processing hierarchy, resulting in prediction errors. In turn, prediction errors are sent back to the higher level, via forward connections, which sharpen the fit of the probabilistic models, approximating their next predictions to the information represented in the lower level.

Comparisons between downward sweeps of prediction and upward sweeps of sensory input lead to a cascade of error signals throughout the hierarchy. These error signals are used to adjust the forward models that are guiding action execution (Clark, 2013). Importantly, actions are aimed at minimizing prediction errors, by matching sensory inputs to predictions—a process termed *active inference* (Friston, 2008; Friston & Frith, 2015a, b). Eventually, active inference leads iteratively to a solution that will take the organism from its current motor state to the desired one.

For the present review, it is important to point out that the Auxiliary and Integrative approaches share three key computational notions: prediction, error minimization, and hierarchical architecture. In both accounts, *sensory predictions* are used to guide motor control. These predictions are crucial to adaptive motor control, because they allow the organism to overcome spatial distortion and temporal delays that would occur if motor control would depend solely in received sensory input (Miall & Wolpert, 1996; Wolpert, Miall, Britain, Senior, & Trust, 1996). *Error minimization*, between the predictions and the received signals, allows cognitive systems to infer solutions to complex problems (Friston, 2003; Friston, Mattout, & Kilner, 2011; Todorov, 2004). *Hierarchical modeling* allows motor control to be implemented both vertically (ranging from high-level symbolic representations to low-level sensory and motor features) and horizontally (multiple probabilistic models are considered in parallel) (Clark, 2013). The purpose of this review is not to adjudicate between the integrative

and auxiliary accounts. Instead, our goal is to highlight the three key processes that are shared by both accounts to discuss joint action.

Applying hierarchical processing to the social domain

Wolpert et al. (2003) were the first to suggest that there might be a computational parallel between motor control and social interaction. The authors posit that the hierarchical-MOSAIC, initially devised to account for individual motor control, also can sustain social operations. According to hierarchical-MOSAIC, the computational mechanisms that support motor control can also support social functions such as *action recognition* (i.e., interpreting the goal of someone else's actions) and *social mimicry* (i.e., reproducing a movement after observing another person perform a similar movement).

We begin by considering the architecture of Wolpert et al.'s (2003) hierarchical-MOSAIC in the context of action recognition. Imagine you (called the *observer* for convenience) are watching someone else (called an *actor* for convenience) reach out to pick up a cup of coffee. How can the observer infer that the actor's reaching movement is triggered by the actor's internal goal of wanting to have a sip of coffee? According to hierarchical-MOSAIC, the modules at different levels of the hierarchy represent different categories or classes of description of the observed action. The lower level sensory modules represent different observable action elements (e.g., arm position, hand configuration). The middle-level modules represent different sequences of those elements (e.g., different possible reach trajectories and hand configurations). The highest level modules represent different, competing goals and intentions (e.g., "pick up the coffee cup" or "grasp the pot of sugar"). As the observed movement unfolds, parallel predictions made by the probabilistic models at each level are compared to incoming sensory signals. As the observed hand and limb approach the coffee cup rather than the pot of sugar, the models that reflect "pick up the coffee cup" provide a better fit to the sensory information than the models that reflect "grasp the pot of sugar." At the highest and perhaps even the middle level of the hierarchy, observers are using their own motor control mechanisms to assist them in understanding an observed actor's goal.

Wolpert et al. (2003) suggest that a similar process can sustain social action mimicry. In this case, two hierarchical-MOSAIC structures would be involved. One is dedicated to the planning and executing of one's own actions. The other is dedicated to processing someone else's actions. Mimicry could occur through an attunement between both hierarchical-MOSAICS at the lower and middle-levels of the hierarchy, even in the absence of attunement at the goal level. Wolpert's et al. (2003) speculate that how successfully one

comes to understand or reproduce another's actions will depend on the similarity of the hierarchical-MOSAIC that generated an actor's behavior and the hierarchical-MOSAIC of the observer that interprets it. Wolpert's et al. (2003) close their discussion by suggesting that developing an understanding of another individual's intentions entails learning hierarchical-MOSAICS for a partner and modeling the differences between their hierarchical-MOSAIC and our own.

It is important to note that Wolpert's et al.'s (2003) goals were more ambitious than simply proposing a model for action recognition and social mimicry. Their larger claim was that a wide spectrum of social interactions follows the same basic principles of hierarchical predictive processing. Yet, this is where their approach faces some challenges. Wolpert et al. (2003) point out what makes social interactions difficult to capture in a computational model. The authors highlight two general difficulties: (i) The overwhelming variety of potential action, including communicative ones. This poses serious problems for the spatial and temporal resolution of the prediction for a given response; (ii) Variability of time delays. The time delays between communicative actions and their social consequences can range from seconds to days. Considerations, such as these, make it difficult to apply the model concretely to a large number of real-world social interactions, where actors do not subscribe to the same interaction end-goal, which often is open-ended and multifaceted.

Variability in joint-actions is constrained by the assumption that both partners are behaving in the service of a mutually agreed upon action goal. Under everyday circumstances, where joint action is typically intentional and evolving in real time, the variability is narrowed down by two additional factors. Cooperators share similar symbolic representations of the goal, promoting interpersonal attunement of high-level processes. They also share the same external environment, promoting interpersonal attunement of low-level processes. This means the circumscribed case of joint-action reduces the complexity of the social interaction to a level at which we can usefully apply the principles of hierarchical processing. Nevertheless, shared goals and shared environments do not eliminate all of the remaining degrees of freedom. For example, each partner's internal representation of the shared goal may deviate somewhat from one another. Furthermore, a shared environment does not equate to entirely similar sensory inputs between the partners. Partners will have different spatial perspectives by virtue of their different spatial locations and orientations in the environment. Moreover, each partner has distinct roles in the joint-action, which also will define their sensorium in specific ways. The proposed idea is not that top-down (i.e., shared goals) and bottom-up (i.e., sensory inputs) signals are entirely overlapping between joint-action partners, rather shared goals and environment in a joint action context can contribute to a reduction in the degrees of freedom by increasing the overlap between top-down and bottom-up

signals. This means that the decision to cooperate reduces the space of possible predictions about the partner's actions.

Recently, Keller, Novembre, and Loehr (2016) presented a conceptual framework that applies Wolpert's sensorimotor notions to musical joint-actions. Ensemble music making, such as a string quartet, requires three types of internal models. (i) *Self-internal models* are responsible for action planning and motor control. They comprise inverse models that output motor commands and efference copies of one's own actions, along with forward models that estimate the next states of one's own motor system. Each member of the quartet will have different self-internal models of the notes played by their own instrument. (ii) *Other-internal models* support the prediction of other's actions. They are achieved either via top-down processing (when an individual has preexistent knowledge of the other's goal) or via bottom-up processing (when predictions are purely driven by incoming perceptual information). Each member of the quartet will have expectations of the notes they hear from the other instruments in the group. (iii) *Joint-internal models* keep a dynamic representation of the shared goal. Self and other internal models are combined to yield a joint predicted state and a joint estimated state. Quartet members also will have expectations about the overall sound that should be produced during the performance.

Note that this proposal aims at accounting for both compensatory and anticipatory control in joint-action. Compensatory control refers to the ability to correct deviations from the shared goal. Keller et al. (2016) suggest that this is done by updating self-inverse models using the error signal obtained from comparing the actual joint state and the joint desired state. Anticipatory control refers to the ability to execute actions based on predictions. Keller et al. (2016) suggest that this is achieved by an adaptation signal that results from the comparison between predicted joint state and the desired joint state.

Keller et al.'s (2016) proposal frames Wolpert's forward model within the joint-action context. However, a closer look at Keller et al.'s proposal reveals three potential shortcomings: (1) *Lack of feedback to shared goal representations*. The uppermost joint goal level is not updated based on information in the lower levels. This limits the model's ability to account for flexible joint-goals (e.g., switching between the shared goal of carrying a table or the bench depending on the location of both objects). (2) *Lack of feedback to the other-internal model*. Whereas predictive joint-error signals are fed back to the self-inverse model (allowing for compensatory and anticipatory action planning), these error signals are not fed back into an inverse model for a joint-action partner. This discrepancy allows for the practical consequence that the self and other models may gradually diverge over time. (3) *Lack of specificity of the feedback signals*. Keller et al.'s proposal does not specify how sensory input is differentially used to update self and other models. This may potentially limit the model's

ability to account for learning and adaptation within joint-action. Taken together, these shortcomings suggest that Keller et al.'s proposal does not completely account for the flexibility and robustness of joint-actions.

Wolpert and Keller's proposals follow an Auxiliary Forward Model account of social interaction. In contrast, a recent proposal by Friston and Frith (2015a, b) presents an Integrative Forward Model approach to communication. Although described in terms of auditory communication, this approach can be considered a general framework for joint-action. The proposal considers communication as emerging from two brains, each modeling the sensory predictions about their own and each other's behaviors. This formulation implies that "brain a" has a model of "brain b," which in turn has a model of "brain a," and so on. The recurrent predictions of both brains about one another are grounded by their computational similarity, i.e., each brain generates predictions of the sensory outcomes caused by itself and the other in the same way. The authors propose that this will lead to generalized synchronization between the internal models of both brains, such that they will be able to accurately predict each other's behavior.

According to this predictive account of communication, listening and speaking states are generated by the same hierarchical predictive mechanism. These states are generated from exteroceptive and proprioceptive prediction errors. In particular, listening is conceptualized as the predictive process through which one attributes the sensory experience of hearing sounds as having been generated by the other. In listening states, exteroceptive prediction errors reflect the differences between our internal model of the other (who is generating the sensations) and the sensations themselves. In contrast, during speaking, exteroceptive information is attenuated, a phenomenon known as sensory attenuation (Brown, Adams, Parees, Edwards, & Friston, 2013) and proprioceptive information gains relevance. Speaking is an instance of active inference and is achieved by enacting the proprioceptive predictions generated by the hierarchical system. The observation that trying to speak and listen at the same time attenuates the exteroceptive signal (Numminen, Salmelin, & Hari, 1999) is evidence of the inherent tradeoff between exteroceptive and proprioceptive signals in hierarchical processing.

Friston and Frith (2015a, b) claim that to successfully take turns between listening and speaking both individuals must share the same higher order belief about when to speak and when to listen. The hierarchical predictive systems of both individuals become coupled due to sharing similar substrates (action-perception mechanisms) and sharing similar top-down expectations (beliefs about when to speak and listen). When this occurs, both exteroceptive and predictive errors are minimized, resulting in successful communicative coordination.

Friston and Frith (2015a, b) also offer a description for coordinated turn-taking during communication. Applying this view to the more general case of joint-action, however, requires the consideration of two issues in greater detail: (1) *Simultaneous self-other actions*. During joint-action, partners often must act simultaneously with one another, instead of taking turns, as is usually the case in verbal communication. For example, carrying a table together requires partners continuously to integrate ongoing predictions about the other's actions in their own actions. (2) *Synergy in joint-action*. If generated separately, the predictions about the sensory consequences of self-action and other-action will deviate from the experienced sensory consequences of the joint-action. For example, the predicted consequences of each individual lifting one side of a table independently will differ from the predicted consequences of both individuals lifting one side of the table simultaneously. Joint-actions often entail synergy between individual actions, and so, to minimize prediction error during joint-actions, predictive mechanisms must generate synergistic predictions. In contrast to the more general case of joint action, where simultaneous action and synergy are ubiquitous, these two features are not considered in the modeling efforts of Friston and Frith's (2015a, b).

In summary, previous forward model proposals do not fully map onto the minimal requirements for a joint-action framework as delineated by Vesper et al. (2010). At the same time, Vesper et al. (2010) do not offer the computational rigor to be found in forward modeling computational proposals (Wolpert's et al., 2003). In the present review, we seek to bridge this gap by offering a hierarchical predictive framework for joint-action, the predictive joint-action model (PJAM). This theoretical framework builds on previous forward model approaches by incorporating the necessary higher-order architecture to deal with the specifics of joint-action implementation (Vesper et al., 2010).

We test the success of this venture in two ways. First, we see to what extent this framework is able to account for the burgeoning empirical literature on joint action research. Second, we test the proposal by seeing to what extent it leads to novel and testable hypotheses about joint action. By engaging in those two exercises, we hope to contribute to the development of new theoretical ideas that can encompass the empirical observations that do not yet fit well within this framework.

Predictive Joint-Action Model (PJAM)

How are two individuals able to coordinate their actions to achieve a shared goal? To help ground this problem, we use the scenario of two young brothers helping their mother prepare a birthday party (Sebanz, Bekkering & Knoblich; 2006). The mother has asked the boys to set-



Fig. 1. Joint-action of two young boys carrying a table down some stairs. Image by James Aldridge retrieved from <http://jamesaldridge-artist.co.uk/blog/?p=346>

up a birthday lunch in the garden. The boys begin by jointly carrying a table from the terrace to the garden's corner (Fig. 1). Although the boys share the same goal, each one must contribute differently to the task. One must lift the back of the table, while the other one supports the weight at the front. The boys must continuously adapt to each other's movements while carrying the table and navigating their way toward their desired destination while avoiding obstacles (e.g., rocks, drying racks, etc). As we will see, this seemingly simple task belies a complex interlocking of intentions, actions, and perceptions between the two boys. In the following sections, we use this example to delineate, level by level, PJAM's hierarchical predictive framework for joint-action.

The predictive joint-action model (PJAM) is conceptualized minimally as three hierarchical processing levels: *goal representation*, *action-planning*, and *sensory routing*. These levels are illustrated in Fig. 2. In keeping with the hierarchical predictive approach (Clark, 2013; Wolpert et al., 2003), each level encodes parallel possible states (probabilistic models) that generate predictions about the states represented on the level below. Continuous comparison between upper-level state predictions and their adjacent lower-level states, leading to subsequent error minimization, allows individuals to find motor solutions that will move them toward their shared goal. Next, we describe the operations occurring at each level of the hierarchy.

The *goal representation level* at the top of the hierarchy models symbolic representations of shared goals. Parallel probable shared goals co-exist at this level (e.g., models that represent carrying the table together, models that represent

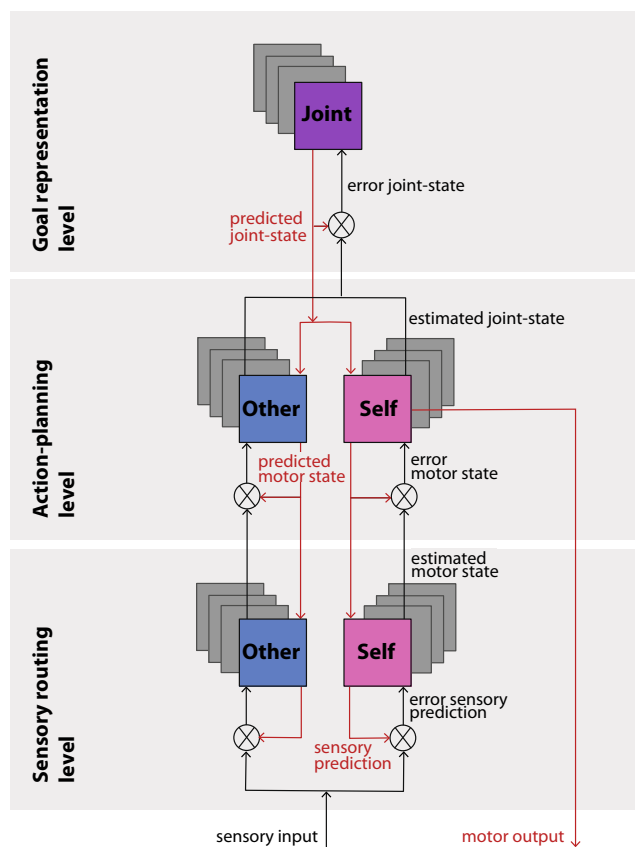


Fig. 2. Predictive joint-action model (PJAM), which is minimally composed of three levels: goal-representation, action-planning, and sensory routing. The framework assumes that each partner in a joint-action maintains internal models of both themselves and their partners. The goal-representation level is responsible for maintaining and updating the shared goals guiding the interaction. The action-planning level outputs motor commands that take into account both the desired contributions of oneself and one's partners to the interaction. The sensory routing level receives the inflow of sensory input and compares it to internal model predictions pertaining to each partner's action outcomes within the interaction. Each level generates predictions of the information that it expects to be found on the level below. Continuous comparison between adjacent levels results in error signals that are sent up to optimize subsequent predictions in the level above.

carrying a bench together, models that represent carrying a tray with plates and glasses, etc.). Symbolic representations can be developed via verbal agreements, cultural conventions, common sense knowledge, and previous experiences (Clark, 1996). Clark and Brennan (1991) propose that joint-action requires interactors to share common ground, i.e., mutual knowledge, mutual beliefs, and mutual assumptions about a possible shared goal and its context. But what does it mean for joint actors to subscribe to a shared goal? Philosophical approaches to joint-action suggest that subscribing to a shared goal implies mutual responsiveness. Both actors are responsive to the intentions and actions of one another, they are both committed to the joint activity (even if for distinct reasons), and they both are committed to mutual support (Bratman, 1992). Notably, Vesper et al.'s (2010) proposal does not

require mutual responsiveness, commitment, or mutual support, even though these features often may be in play.

In our example, the boys have verbally agreed on the goal of cooperating with one another to carry the table from the terrace to the yard. Therefore, the cluster of models that represent the shared goal of carrying the table together will be activated. The processing hierarchy treats the shared goal as an actual state of affairs, causing a cascading downward prediction of what should be experienced next in the levels' below. The goal representation level will output an abstract representation of the next desired joint-state to the level below (e.g., a symbolic representation of moving the table together) and receive as input information about the currently estimated joint-state provided by the level below (e.g., representation of the interdependent states of the cooperators). The continuous comparison between these two sources of information will generate a signal indexing the deviation between the actual state of the joint system and its desired state (i.e., joint-state error). This error signal is fed back to the goal representation level, where it is used to update and select the generative models responsible for outputting predictions about the next desired joint-state. Thus at each iteration, the output of these models will become more effective in guiding action-selection that will lead the boys to get closer to their goal.

At the *action-planning level*, the joint-state models encode each individual's expected contribution to the desired joint-state. A given desired joint-state can be achieved through several different configurations. For example, to transport the table from one spot to another, the older brother can hold the front of the table and the younger brother the back, or vice-versa. The brothers can both face forward or they can face each other. The brothers can grasp the wider or narrower side of the table, with one brother taking more of the load than the other. PJAM proposes that models of the self and models of partners can be paired to represent the possible combinations of individual contributions to the joint action. These models transform the desired joint-state signal, descending from the goal representation level, into expectations of the unique motor states of the self and the partner. Predictions about self and other motor states are iteratively compared with estimated motor states that arrive from the layer below, and the resultant discrepancy is the motor state prediction error. These errors help update and select the self-other models that provide a better fit for the motor states estimated from the lower level.

In the table-carrying task, more than one pair of self and other model will likely be compatible with the shared goal. Parallel simulations of these motor plans, along with comparisons of their prediction errors, would help determine which model results in the least effort. At each moment in the joint-action, the self-model integrated into the activated self-other model will output a command for action execution. This output corresponds to the model's prediction of one's own next

motor state. Because self and other models are complementary, this motor command inherently integrates the internal predictions about one's partner's next motor states. In this way, partners are able to bypass the temporal delays that would otherwise arise from waiting to receive sensory information about each other's executed actions before being able to plan and execute their own actions.

The process outlined above has two inter-related strengths. First, it filters out self-other models that have a poor fit between the desired joint-state (top-down information) and the individual motor estimates (bottom-up information). Second, the iterative, real-time nature of the process means that with every update of new sensory information, a candidate self-other model can be deactivated in favor of one that produces a better fit.

At the *sensory routing level*, the incoming sensory information reflecting the outcome of the joint-action is compared to independent sensory predictions of one's own and of the partner's expected actions. This comparison serves as a gate for parsing one's own and the partner's sensory information into their corresponding predictive streams. Deviations between sensory input and sensory predictions (i.e., sensory predictive errors) are fed-back to sensory predictive models in order to continuously improve sensory parsing. This allows the predictive system to attribute external consequences to each individual's actions. For example, each brother will receive haptic input from the table. This input will be confounded with respect to its source, in that the haptic input itself does not differentiate between the forces that each brother applies to the table. However, the comparison between the haptic input and the separate predictions about one's own and the other's action will help feed each predictive cascade of self and other into their respective streams.

In summary, the top level in PJAM represents a symbolic joint task, which is bifurcated at both the lower *action-planning layer* and *sensory routing layer*. We recognize that these three-levels are offered here as a theoretical proposal in fairly broad strokes. To implement fully the joint-actions computationally, each of the three levels would necessarily contain finer-grained local processing hierarchies dedicated to representing different aspects of the joint-action. For example, inside the action-planning level, several sets of local hierarchies need to be dedicated to translating low-level movement primitives (e.g., groups of muscles) into middle-level combinations of movement patterns and trajectories.

Theoretical model assumptions and empirical evidence

PJAM's architecture makes explicit several important assumptions. Some of these assumptions have support in the current empirical literature, whereas others remain to be

tested. In this section we first demonstrate the usefulness of PJAM as an organizing framework by comparing its key assumptions with existing evidence. We then demonstrate the generative utility of PJAM by pointing to areas where new studies are needed to provide important detail.

Goal representation level

PJAM's *goal representation level* implies that individuals engaged in joint action encode deviations between joint-state estimates and shared goal representations. This assumption underlies successful performance in ensemble music. When musicians play together they must maintain a representation of the desired unified sound of the ensemble (Keller, 2008). This observation is supported by evidence from two relevant musical ensemble studies. Loehr et al. (2013) showed that during ensemble playing, musicians kept track of deviations from the joint-desired sound, and Loehr and Vesper (2016) showed that playing together leads to the formation of shared expectations about the desired ensemble sound.

Loehr, Kourtis, Vesper, Sebanz, and Knoblich (2013) tested joint expectations by recording event-related potentials (ERP) from the scalp of each member of a piano duet while they played a previously memorized duet. The goal was to see whether pianists monitored unexpected deviations from the shared sound goal. To do so, the experimenters programmed some of the piano keys to generate altered pitches that either did or did not change the joint auditory outcome (i.e., the harmony of a chord resulting from the two pianists' combined pitches). The feedback-related negativity in ERP signals, which are triggered by unexpected events, were now observed in response to alternations in the joint chord.

In Loehr and Vesper (2016), pairs of participants learned to play a melody together. After learning it, they were asked to play the melody while hearing only themselves or hearing both partners. The results showed that participants made fewer errors when hearing both partners. This finding suggests that when learning a novel joint-action, partners form representations of the shared goal and use this shared representation to guide future actions. These findings support the following PJAM features at the *goal representation level*: (i) Partners in a joint-action have the capacity to monitor shared states against the desired shared goal; (ii) Prediction errors in joint-action play a functional role, because deviations from the expected joint-action states are indexed by feedback-related negativity ERPs, known to encode unexpected events. Yet empirical studies addressing the mechanisms involved in monitoring joint-states against the symbolic representations of shared goals are still scant and, therefore, remains an open-question in the joint-action literature (Keller, 2008; Sebanz & Knoblich, 2009).

Action planning level

Action planning necessarily involves coordination between several levels in a hierarchy, spanning from high-level planning (e.g., whether to face the partner and move backwards or not face the partner and move forward), to mid-level planning (e.g., estimating waypoints for movement trajectories), and to low-level planning (e.g., activating specific groups of muscles). These processes in individual action planning have been described in detail elsewhere (Haruno et al., 2003) and are not the focus of this paper. We focus on how two sets of individual action planning models (self and other), which are each assumed to include the necessary high to low-level local hierarchies entailed in individual motor control (Haruno et al., 2003), are jointly generated and updated to minimize prediction error and lead to successful joint-actions.

PJAM posits that the action-planning level maintains probabilistic models of the expected motor contributors from each individual to the joint action. The organization of PJAM's action planning level has five separate implications: (i) We keep internal models of both our own and our partners expected actions in a joint-task; (ii) We generate predictions about the future actions of self and other in a joint-action; (iii) Predictions about a partner's actions are contingent on the predictions of our own actions; (iv) We encode motor prediction errors about our joint-action partners; (v) Successful joint action depends on accurate models of our partners. Next, we discuss evidence that relates to each of these assumptions.

(i) *We keep internal models of both our own and our partners expected actions in a joint-task*

At the *action-planning level*, pairs of models (our own and our partner's motor states) represent the expected contributions to a joint-action. There is extensive evidence in the literature that individuals maintain and update both types of internal models. Several studies indicate that individuals spontaneously and involuntarily maintain representations of each other's tasks when asked to share responsibility for a task (Atmaca, Sebanz, Prinz, & Knoblich, 2008; Sebanz, Bekkering, & Knoblich, 2006; Sebanz, Knoblich, & Prinz, 2003, 2005). A popular method used to study this question is the social Simon task (Sebanz, Knoblich, & Prinz, 2003). This method compares individuals' performance in the well-known Simon task when performed alone and bimanually (Simon, 1969) with performance when each participant is responsible for only one of the two responses (social Simon task).

In the standard Simon task, participants respond bimanually by pressing a left or right button in response to stimulus features such as pitch or color (Simon & Craft, 1970; Richard Simon, 1969). Although stimulus location is not relevant to the task, in a color task blue and red stimuli are presented

randomly on the left or right side. Results show that responses are faster when stimulus and response are spatially compatible, whereas noncorresponding stimulus-response pairings result in slower responses. The theoretical interpretation is that a spatial mismatch between stimulus and response (even when space is irrelevant to the task) creates a competition between the required response and the primed response (Kornblum, Hasbroucq, & Osman, 1990). Notably, if the stimulus-response feature overlap is removed, by presenting a task with only one response (i.e., a go/no-go task), the effect disappears (Liepelt & Prinz, 2011). This pattern of results is known as the Simon effect.

In the social Simon task, responses are shared between two participants, so that each individual presses only one of the two keys (right or left) in response to their assigned stimulus (e.g., red or blue cue). Thus, an ideal strategy would be for participants to ignore each other's part of the task. Surprisingly, most participants are unable to do so. The main finding is that the individual Simon task and the social Simon task result in comparable performance, which both show more interference than in the individual go/no-go task. The standard explanation is that when participants are representing their and their partner's actions, the stimulus-response conflict is reinstated even for a single stimulus-response mapping (Knoblich & Sebanz, 2006; Sebanz et al., 2003).

Converging evidence for task co-representation comes from adapting other classical stimuli-response compatibility tasks for the social realm. For example, the joint flanker effect demonstrates that co-representation is not restricted to tasks initiating spatial interference, but generalizes to tasks involving arbitrary stimulus-response associations (Atmaca, Sebanz, & Knoblich, 2011). Additionally, a compatibility effect between numerical and spatial stimuli known as the SNARC effect has been observed in the social transformation of this task-joint SNARC effect (Atmaca et al., 2008).

The theoretical interpretation in terms of shared representation is not held universally. Recent data suggest that a social dimension is not even necessary. For example, Dolk, Hommel, Prinz and Liepelt (2013) substituted one of the participants in the social Simon task for a Chinese cat toy with a moving arm. The authors showed that if the toy provided a strong spatial reference, participants allocated task co-representations to this inanimate object. This conception is in line with the view that social and non-social events are processed in similar ways, though social events often are more salient, recruiting more cognitive resources (Friesen & Kingstone, 1998; Langton & Bruce, 2000).

Other findings point to dynamic interactions between situational variables and the social Simon effect and thus to the conclusion that modeling the other's task is obligatory. For example, a participant's mood can mediate the effect (Kuhbandner, Pekrun, & Maier, 2010). Participants in this study watched a positive, negative, or neutral film clip before

performing the task. Although the expected social Simon effect was observed in the neutral mood, under a positive mood it was even stronger, and under a negative mood the effect was absent. This finding is congruent with other observations that individuals in a positive mood find a stranger to be more likeable (Baron, 1987) and show greater tendency toward cooperation (Forgas, 1998).

A participant's concept of self also can play a role. Colzato, de Bruijn, and Hommel (2012) primed participants with either an individualistic or a social orientation before completing a social Simon task. Results showed the effect to be stronger in the socially primed group. In a related study (Colzato, de Bruijn, et al., 2012), pairs of Buddhists and pairs of atheists performed the social Simon task; Buddhists yielded the larger effect. The authors interpreted this as consistent with the Buddhist emphasis on self-other integration.

Whether participants perceive themselves to be cooperating or competing also has an influence on the social Simon effect. In Iani et al. (2011), a social Simon effect was observed when participants were in a cooperative condition, but it disappeared in the competitive condition. In de Bruijn, Miedl, and Bekkering (2008) participants were asked to respond faster than their competitor. No-go trials were of two different types: In compatible no-go trials both participants should refrain from responding, whereas in incompatible no-go trials only one of the participants should not respond. The prediction was that, if automatic task co-representations were formed, anticipating the partner's part of the task would lead to slower responses in incompatible trials. The main finding was that only the participants who were not successful in competition were impeded by incompatible actions. All of these findings point to complex interactions between a participant's intent and situational demands in determining whether a co-actor's actions are modeled or ignored.

It is important to caution that the tasks used in these studies (e.g., social Simon, joint SNARC) do not have all the requirements of joint-action, as defined by Sebanz, Knoblich, et al. (2006). In particular, there is no joint goal that requires participants to coordinate with one another to achieve it. This means that these tasks cannot be used to test whether or not we keep internal representations of self and other's contributions to the performance of joint-actions. In one study offering preliminary support for this idea, the experimenter asked pairs of participants to jump next to each other with the goal of landing at the same time (Vesper et al., 2013). Crucially, the jump distance was not always the same for each partner. Furthermore, the participants could not see each other and were asked not to communicate with one another, allowing the experimenters to control the available information about the partner's jump. Participants were permitted to see the distance that they and their partner would need to jump on the upcoming trial. Furthermore, a tone signaled when they and their partner landed. Thus, to satisfy the joint goal, participants

had to plan their jumps based on knowledge of the previous jump distances and tones signaling the timing of each partner's landing. The results showed that participants adjusted their jumps to the specific difference in distance between each participant's jumps, reducing the discrepancy in timing over iterative jumps. Vesper et al. concluded that participants rely on motor simulations to achieve the goal of landing simultaneously and that the coordination effort is distributed based on the task characteristics. Overall, these findings are consistent with the notion that partners in a joint-action maintain representations of both one's own and one's partners expected contributions to the task, i.e., self-other models.

PJAM assumes that self-other models are encoded at the motor level. In agreement with this assumption, a neural signature of motor preparation, the lateralized readiness potential, emerges during turn-taking joint-tasks regardless of whose turn it was to respond (Holländer, Jung, and Prinz (2011). This observation is consistent with the notion that each partner maintains covert motor activations relating to the expected contributions of their partner in a joint-task. There is evidence that these activations usually remain covert because of neural inhibition (Sebanz, Knoblich, et al., 2006; Tsai, Kuo, Jing, Hung, & Tzeng, 2006). Nevertheless, there are times when these activations lead to overt action in the form of involuntary compensatory movements (Sebanz & Shiffar, 2007). For example, Häberle et al. (2008) reported that individuals who were not directly involved in a cooperative balancing task, but were observing their teammates perform the task, tended to make small movements congruent with the achievement of the balancing task. This is evidence that partners encode each other's contributions to the shared task at the motor level.

PJAM assumes that self-other internal models generate motor predictions about each contributor to the joint-action. These predictions are then compared to estimated states resulting in prediction errors that ultimately contribute to the success of the joint-action. Following this reasoning, the extent to which we are able to encode our joint-action partners at the motor level should have an impact on our ability to successfully cooperate. Accordingly, Novembre, Ticini, Schütz-Bosbach, and Keller (2012) showed that successful temporal coordination during a musical duet was positively related to the extent to which musicians encode each other's motor actions, as measured by corticospinal activation after transcranial magnetic stimulation to the right primary motor cortex.

In PJAM, internal self and other models are activated recursively through top-down and bottom-up interactions. This leads to the activation of the representations that are relevant to the achievement of a given shared goal over competing representations that are less relevant. This process is constrained, because the mere presence of a social partner in a cooperative setting is not sufficient to initiate it. As recent studies show, the extent to which one encodes other's contributions to a joint-action is influenced by contextual factors.

For example, Iani, Anelli, Nicoletti, Arcuri, and Rubichi (2011) observed that the social Simon effect occurred when partners shared the goal of maximizing team performance but vanished when their goals were to maximize individual success. This suggests that the social Simon effect is mediated by the existence of shared goals. Further support for this idea is given by the observation that anticipatory ERPs are greater when anticipating the joint-action of partners than when anticipating the action of bystanders (Kourtis, Sebanz, & Knoblich, 2010).

The claim that encoding a partner's actions is a self-referential process raises the question of whether this process is permeable to individual differences. Are some individuals more prone to represent their partners than others? A positive answer to this question is offered by Novembre's et al. (2012) observation that self-reported empathy is positively correlated with the strength of a partner's motor contribution to the joint-action. Similarly, de Bruijn, Miedl, and Bekkering (2008) report individual differences in the degree to which a partner's task is represented in a social Simon task. Whereas some participants tended to slow down their responses to encode their partner's side of the task, others did not. Importantly, these individual differences have been shown to be sensitive to experimental manipulation and long-term training (Colzato, Zech, et al., 2012; Colzato, de Bruijn, et al., 2012). This evidence is consistent with PJAM, which proposes that context-specific internal models about a partner are updated continuously, leading to learning and improved encoding of the partner's contributions to the task.

(ii) *We generate predictions about the future actions of self and other in a joint-action*

PJAM proposes that predictions of our partner's action are continuously compared with incoming sensory information about it. This implies that observing another person is an active process. This idea was supported in a study where participants either stacked blocks or watched another participant stack blocks while their eye movements were tracked (Flanagan & Johansson, 2003). Normally, when stacking blocks, our eyes move to the block's intended location in advance of the hand. The authors found that this pattern of eye movements in participants regardless of whether or not they stacked the blocks themselves or watched someone else do it (Flanagan & Johansson, 2003; Gredebäck & Falck-Ytter, 2015).

These predictions are generated by simulations about someone else's goal-directed actions (Sebanz & Knoblich, 2009; Wilson & Knoblich, 2005). Some evidence suggests that these simulations exploit the natural statistics of human biomechanics (Graf et al., 2007; Parkinson, Springer, & Prinz, 2012; Sparenberg et al., 2012). For example, Graf et al. (2007) presented participants with point-light displays of individuals

walking towards and then behind a screen. After variable time intervals, the walker emerged from behind the screen as a static point-light display. Participants were asked to decide whether or not the static posture was a continuation of the walking sequence. The results showed that identification accuracy improved when the occlusion time matched the time necessary for the walker to reach the static position shown at the end. This finding supports the idea that predictions are derived from simulations about other's actions that are informed by prior expectations learned in the environment.

Despite a wealth of evidence suggesting that action observation and simulation entails activation in ventral premotor cortex (for review, see Rizzolatti, Fogassi, & Gallese, 2002), exactly what this activation reflects is open to a range of interpretations. One possibility is that it reflects a prediction about the future motor state of the partner. According to PJAM, the motor system is necessary for such prediction to be made. Another possibility is that the prediction is entirely perceptual in origin. In short, the motor system may not be necessary at all. A third possibility is that the activation does not reflect a prediction at all. It could reflect a more general semantic description of the images

(iii) *Predictions about a partner's actions are contingent on the predictions of our own actions*

PJAM posits that in joint-action settings, predictive models about one's partner are inherently linked to predictive models about oneself, i.e., through self-other models. This means that expectations about the joint-task motor contributions of self and other are interdependent and contingent on the ongoing interaction. Support for this assumption can be found in a recent study showing that participants who learned to perform a specific task in a joint context, were later better able to perform that task together than individually. This finding implies that participants learned not only their side of the task, but they also learned how the other's performance affects their own, resulting in the observation of diminished individual performance after joint learning (Loehr & Vesper, 2016).

Additional support for the idea that we simulate our partner is found in reports that individuals not only encode expectations about the general nature of other's actions but also encode expectations about detailed temporal and spatial features of them. For example, Neri, Luu, and Levi (2006) asked participants to identify a target partner in a noisy point-light display in either a dancing or a fighting action sequences. They reported that participants had more difficulty identifying their partners when the point lights were desynchronized rather than synchronized.

This finding has been extended to communicative interactions. Manera, Schouten, Verfaillie, and Becchio (2013) used point-light displays to depict scenes in which one agent either pointed towards the ground or simply turned around after

which the other agent would squat. The authors varied the timing between the actions of the two agents. Participants were more accurate at detecting the target partners when the interaction maintained its ecological timing. These studies suggest that representations about each individual in a social interaction are encoded within the joint context of the interaction. It is important to note that these studies do not give us direct evidence of the interdependence of self and other representations, but they do show that we register interdependency between two observed interactors. Further studies are necessary to test the interdependence of self and other representations during joint-actions.

(i) *We encode prediction errors about our partner's motor states.*

In the previous section, we discussed Loehr et al. (2013) in the context of shared goal representations. The same study adds to what is known about prediction errors in joint-action at other levels. The results showed that during joint musical activity, each performer tracks deviations from expectations about the self, about the other, and about the joint outcome. The findings suggest further that these different types of prediction errors are prioritized in different ways. Using the P300 ERP component as a marker of attentional importance, the authors report that prediction errors related to the joint-goal were more salient than those related to one's own actions, which in turn were more salient than prediction errors related to the actions of others. These results align well with the PJAM assumption that prediction errors are key components of joint-action processing, both at the individual and at the joint level.

PJAM proposes that we monitor deviations (prediction errors) between the expected states and the actual motor states for both ourselves and our partners in the joint-action. Some research suggests that these processes are supported by partially overlapping neural resources (van Schie, Mars, Coles, & Bekkering, 2004). Using an fMRI task, Radke, de Lange, Ullsperger & de Bruijn (2011) showed increased activation of the medial frontal cortex (pmFC) and bilateral insula, both when making errors that affected one's own economic payoffs and when making errors that affected the partner. In contrast, the medial prefrontal cortex (mpFC), an important part of motor simulation, was specifically activated by errors that affect only a partner's payoffs. This suggests that partners in an interaction process their own errors differently than those of their partner.

The extent to which we monitor other's action failures in a joint-action task is modulated by contextual and personal factors. Some evidence suggests that the monitoring is sensitive to the importance of the other's errors to our own goals. An EEG study by Koban, Pourtois, Vocat, and Vuilleumier (2010) showed that the amplitude of error-related negativity

potentials (ERN) was stronger when individuals detected mistakes made by their partners than when they detected mistakes made by independent competitors. The salience of self-referential prediction errors is further supported by reports that the amplitude of ERNs are larger when participants are observing people of close social affiliation, either through long-term friendship or short-term social priming (Carp, Halenar, Quandt, Sklar & Compton, 2009; Kang, Hirsh & Chasteen 2010).

(ii) *Successful joint action depends on accurate models of a partner.*

PJAM is premised on an error minimization strategy, meaning that prediction errors are used to improve action outcomes. It is commonly observed that we get better at cooperating with others the more we experience it. According to PJAM, predictive expertise can occur at two inter-related levels: (1) expertise in predicting the consequences of one's own actions within the joint-action (minimizing error in self-predictive models), and (2) expertise in predicting the consequences of others actions within the joint-action (minimizing errors in other-predictive models).

Loehr and Vesper (2016) reported that expertise in predicting the consequences of one's own actions is dependent on the learning context. In this study, pairs of participants learned how to play a melody together. After learning to play at a high level of proficiency, participants were asked to play the melody alone or together with the partner. The results showed that participants made fewer errors when playing with their partner. This finding suggests that expertise about one's own actions is contingent on the joint-action context.

Two complementary studies support the idea that expertise in action coordination results in more accurate internal models of other people. First, training improves prediction, which is evidenced by the observation that experts need less information to make accurate predictions and are proficient in anticipating other's errors (Savelsbergh, Williams, Van der Kamp, & Ward, 2002). Second, it is easier to coordinate with oneself than it is to coordinate with another person, probably because our self-models are more accurate than our models of others (Keller et al., 2007). In PJAM, these two types of expertise (i.e., prediction of self and the other) are inherently linked by the bidirectional flow of expectations arriving from the upper shared goal level and lower-level inputs from the sensory routing level. In PJAM, expertise refers to overall error minimization across the predictive hierarchy.

The acquisition of predictive expertise from sensory feedback about a partner's action is crucial when the joint action requires anticipatory control. This is illustrated by Knoblich and Jordan (2003), who devised a task in which participants tracked a moving dot with an overlapping circle. The dot moved horizontally across the screen in one direction until it

bounced off a visible border, immediately reversing its direction while maintaining the same speed. The participants controlled the position of the tracking circle indirectly, by manipulating its speed and direction via the left or right button press. Critically, unlike the target dot, the tracking circle did not automatically bounce off of the border. Thus, the participants had to learn to reverse the tracker's direction through key presses in order to mimic the dot's reversals. As the authors reasoned, minimizing the distance between the dot and the circle meant that participants would need to anticipate the dot's reversal at the border region and learn the number and timing of key presses that would best mimic it. The participants were tested both individually and in pairs. In the individual condition, both keys were under the control of a single participant. In the paired condition, one participant was responsible for the left button while another participant was responsible for the right button. Notably, the members of each pair were separated from one another and, therefore, communication was suppressed. Although performance was by in large better in the individual condition than in the paired condition, this difference was reduced when a tone signaled the partner's button presses. This finding stresses the importance that sensory feedback about others' actions plays in the ability to anticipate another's actions. According to PJAM, incoming sensory information such as this helps to minimize prediction errors along the hierarchy.

Interestingly, partners often adopt strategies that make each member easier to predict to the other, in essence, minimizing each other's prediction errors. Partners will sometimes exaggerate their movements (Goebel & Palmer, 2009) and modulate the temporal variability of their actions (Vesper, van der Wel, Knoblich, & Sebanz, 2011; Vesper et al., 2013) to simplify the required predictions. Vesper et al. (2010) refers to these tactics as coordination smoothers.

The existence of multiple coordination smoothers raises other questions about optimal control in joint-action. PJAM proposes that optimal motor control emerges from iteratively selecting the internal models that best approximate the action goal with minimal physical effort (Wolpert & Miall, 1996). But this implies at least two different metrics of effort. In joint action, both the physical effort required to perform the task and the coordination effort needed to predict one's partner must presumably be weighted to achieve optimal social cooperation. Coordination smoothers therefore provide the opportunity to place physical effort and coordination effort in a trading relationship. For example, if one strategy is to add movement modulation to aid in the making of partner predictions, it has a cost of increasing the physical effort involved. The benefit, however, will be that the extra actions facilitate interpersonal prediction, thus decreasing coordination effort. Future studies are necessary to better understand the trade-offs between physical and coordination effort in joint-action.

Sensory routing level

PJAM assumes that incoming sensory input is continuously compared with downwards predictions about the sensory consequences of one's own and one's partner's actions. These comparisons result in the computation of prediction errors, which are then sent up the self or other action planning level. This architecture also means that the reliability of the sensory input used to compute prediction errors is crucial for optimal processing. The successful routing of the action consequences of self and other will improve the specificity of prediction errors and ultimately improve interpersonal coordination.

Support for the notion that sensory predictions about oneself and one's partners are compared in parallel to the incoming sensory input comes from studies observing interpersonal sensory cancellation. Sensory cancellation occurs when predictions about the motor outcomes of an ongoing action are used to filter-out the sensory feedback produced by that same action (Blakemore, Frith, & Wolpert, 1999). One popular example of sensory cancellation is the observation that it is hard, if not impossible, to tickle oneself. In this case, the prediction of the sensory consequences of tickling dampens the sensory experience of the tickling itself (Blakemore, Wolpert, & Frith, 2000). Interpersonal sensory cancellation occurs in an analogous way when our predictions of someone else's actions dampen the sensorial experience of these outcomes. Sato (2008) presented participants with an auditory tone under three conditions: after the participants pressed a key themselves, after they observed another person press the key, or unexpectedly. Participants perceived the tone as being less intense whenever they or the other person pressed the key compared with the unexpected condition. The authors reasoned that this sensory attenuation occurred, because participants maintain sensory attenuating predictions about the outcomes of both their own and other people's actions.

A recent study reports that attributing sensory input to a partner's actions can result in improved cooperation. For example, Chaminade, Marchant, Kilner, and Frith (2012) performed an fMRI study in which two participants cooperated to jointly control an object presented on a screen using a joystick. Joint-action errors were indexed by computing the distance between the ideal target and the participant-controlled graphic objects. The results showed that the number of errors made during cooperation was negatively correlated with activity in the parietal operculum. This brain region has previously been associated with the integration of the somatosensory consequences of actions. In particular, Blakemore, Goodbody, and Wolpert (1998) reported that this region was activated when participants experienced sensory feedback that did not pertain to their own action. The observation that increased activity in this area was related to a decrease in cooperative error supports PJAM's assumption that attributing

sensory consequences to joint-action partners is linked to cooperative success.

The successful matching of action predictions with sensory outcomes is thought by some to correspond to the experience of agency, which refers to the perception of causing something to happen by intention (Obhi, 2012; Pacherie, 2012; Schüür & Haggard, 2012). However, recent theoretical views suggest that there are at least two levels to the experience of agency that are somewhat dissociable. First, there are lower level processes such as interpersonal action prediction and sensory outcome matching (Blakemore & Frith, 2003; Frith, Blakemore, & Wolpert, 2000). These effects are dissociable from the conscious feeling of agency and explicit attributions of agency to an event (Wegner, Sparrow, & Winerman, 2004). Evidence for this differentiation is presented in a study by Obhi and Hall (2011). The authors used self-report measurements to assess the subjective sense of agency and measured intentional binding to index lower-level processes. Intentional binding refers to the perceived compression in time between a voluntary action to produce an effect (e.g., pressing a key to generate an auditory tone) and the experience of the effect (e.g., perception of the auditory tone). Participants were paired for a joint-action task in which one person pressed a key to generate a tone, and the other person subsequently pressed the key. The results showed that although only the initiator reported reliable subjective sense of agency, both the initiator and the follower demonstrated the same levels of intentional binding. This pattern of results supports the idea that partners process the link between each other's actions and sensory outcomes (lower-level processing) but further suggests that this process alone does not lead to the subjective feeling of agency (higher-level processing).

The routing of joint sensory outcomes into separate predictive hierarchies for self and other is proposed to occur at lower-levels in PJAM's architecture. As such, these processes will not necessarily percolate into conscious experience. Nonetheless, a conscious sense of agency could be conceptualized within PJAM as arising from a perceived match between intentional action planning, encoded in top-down action predictions, and the corresponding sensory effects, encoded in bottom-up processing. However, one piece of evidence contradicts this assumption. van der Wel, Sebanz, and Knoblich (2012) asked pairs of participants to either collaborate in moving a pole back and forth, or do the same task alone. The authors observed that during the first trials, before any learning could occur, the self-reported sense of agency in the joint-task was virtually the same as the self-reported sense of agency in the individual task. The authors interpreted this result as being incompatible with a predictive sensorimotor approach to the sense of agency. Following their line of reasoning, in an individual action task there is a 1-to-1 match between action prediction and action outcome. In contrast, in the joint-action task, the individual action prediction is matched to the

combined outcome of both partner's actions. It could therefore never be as strong a match in the joint task. We propose that reframing van der Wel, Sebanz, and Knoblich (2012) results within PJAM restores the viability of sensorimotor predictive approach to the sense of agency in joint-actions. This is because PJAM proposes that we compare in parallel the predicted effects of our own actions and our partner's actions. Thus, in simple tasks, such as in van der Wel et al. (2012), it is reasonable to assume that the sensory input would be routed in parallel for the action outcomes of each partner. This means that the signal percolating up the self-other hierarchy would pertain specifically to one's intended contribution to the joint-action, leading to a stable sense of agency.

In summary, PJAM assumes that the sensory flow of information during joint-action is routed through both self and other internal predictive cascades. This claim is central to PJAM, because the model relies on the forward sweep of sensory information to train predictions about oneself and one's partners. The sense of agency can be incorporated into PJAM as a consequence of the interaction of top-down and bottom-up processing. It is reasonable to assume that a secure sense of agency is instrumental for successful joint-action, because it allows for conscious top-down modulations of how a shared goal may be broken down into individual contributions. However, the foregoing evidence that lower and higher processes linked to the sense of agency are dissociable highlights the need for further empirical investigation of the topic.

PJAM Assessment

In this section, we consider whether the hierarchical predictive approach proposed by PJAM meets the minimal requirements for a joint-action architecture defined by Vesper et al. (2010). In short, we believe it does. We then discuss how PJAM's organization differs from a direct application of Wolpert, Doya, and Kawato's (2003) proposal to the topic of joint-action, previously unaddressed by those authors. Finally, we consider future avenues for joint-action research derived from PJAM.

Minimal requirements for a joint-action architecture

We have presented PJAM as an application of generic hierarchical predictive processing principles to the specific case of joint-action. The overall success of this endeavor can be judged by asking whether PJAM meets the minimum requirements for an architecture of joint-action as proposed by Vesper et al. (2010):

- (i) *The architecture must support shared goal and corresponding individual task representations.* PJAM proposes that shared goals are represented in a probabilistic fashion at the highest-level level of the hierarchy: *goal representation level*. In the level immediately below, the *action-planning level*, the framework comprises parallel probabilistic models of both one's own and partner's motor contributions to the desired joint-state. Comparisons between these adjacent levels lead to the continuous actualization of shared goal representations.
 - (ii) *The architecture must support the processes of monitoring and predicting each partner's actions.* PJAM comprises parallel predictive cascades for each partner in a shared joint-action. The core cognitive process of this model is *prediction*, which is posited to occur at different levels of abstraction (i.e. from goal representation, to motor planning and sensory expectations) for each partner in the interaction.
 - (iii) *The architecture must allow for continuous coordination.* Continuous interpersonal adaptation is integral to PJAM's organization. The model proposes that the shared goal is broken down into interdependent self-other motor models. These models generate predictions about each individual's contributions to the goal. Self-other models predictions are continuously weighted by incoming sensory information. This process allows for the optimization of the internal representations of each partner in the interaction, supporting the generation of individual motor commands that can account for partner's actions.
- (3) Combinatorial modularity. Internal models about the self and other can be aligned in a vast number of combinations to form self-other configurations. These configurations correspond to the different ways in which each partner's motor abilities might align with one another to approximate them to the shared goal.
 - (4) Signal processing flexibility. The architecture offers flexibility because both top-down and bottom-up streams of information are operating simultaneously, allowing the possible unreliability of each stream to be compensated for by the other. For example, tasks in which sensory feedback information is poor can rely more heavily on top-down signals, derived from previous experience with the task.

Theoretical model specificity

Although PJAM is fully compatible with Wolpert et al.'s (2003) computational proposal, PJAM departs from a direct extrapolation of this framework to the specific case of joint-action in one important aspect. This departure concerns the interdependence of self and other predictive processes. Wolpert et al.'s (2003) framework to joint-action would likely feature two independent hierarchical predictive cascades (i.e., two H-MOSAICs). One cascade would be dedicated to predict the partner's behavior, and another would be dedicated to predict one's own behavior. Each one of these cascades would have their own high- to low-level hierarchy of predictive models dedicated at independently processing each partner's actions. Construed in this way, the success of joint action would depend on the degree to which each hierarchy would be able to account for the individual contributions to the joint action.

In this regard, Keller et al.'s (2016) proposal has an advantage over that of Wolpert et al. (2003). These authors propose that both self and other predictive models are informed by top-down expectations about the joint-desired state. Yet, they do not go so far as to propose that self and other models can be updated in a contingent fashion. Rather, they propose that predictive joint-error signals are fed back to the self-inverse model (allowing for compensatory and anticipatory action planning), but they do not feed these same error signals back into the other-inverse model. This discrepancy opens the door for the self and other models to gradually diverge over time rather than to come together.

In PJAM, the predictive cascades pertaining to one's own and partner's actions are inherently interdependent. A single higher-level layer (i.e., shared-goal representations) informs the generation of two complementary predictive streams about

We believe PJAM meets the minimal requirements for an architecture of joint-action as defined by Vesper et al. (2010). PJAM also is compatible with the Wolpert et al.'s (2003) sensorimotor framework for social interactions, and importantly, extends it to the specific case of joint-action. As such, PJAM is able to cope with the large number of degrees of freedom, both temporal and spatial, that are involved in many social interactions (Wolpert et al., 2003). This is because hierarchical predictive processing supports:

- (1) Specialization of purview and process. For example, in the table carrying task, each actor will have a module dedicated to signaling the haptic forces from the table that are generated by their own actions and another model dedicated to signaling the forces caused by the actions of the other.
- (2) Parallel processing. Candidate modules at each level can be evaluated in parallel, rather than serially, which reduces processing times. For example, one can evaluate multiple self-other configurations to select the one that most closely matches the desired joint-state.

one's own and partner's actions (i.e., self and other models at the action planning and sensory routing layers). In contrast to Keller et al.'s (2016) proposal, the hierarchical organization of PJAM posits that both self and other models are updated by a continuous stream of prediction error signals. PJAM's organization aims to prevent self and other predictive mechanisms to independently diverge from the desired shared-goal. In addition, it facilitates the computation of the one's own contribution to the joint-action, by being contingent on the partner's contributions.

Are the self and other predictive cascades largely independent, as we infer from Wolpert et al., (2003) and Keller et al. (2016), or interdependent as proposed here? Some of the evidence that we have reviewed supports the notion that self and other processing modules are interdependent. For example, Loehr et al. (2013) reported that musicians were sensitive not only to deviations of the self and other from the desired sound (as would be expected from an independent organization) but that they also were sensitive to deviations from the joint product in the desired sound. Manera et al., 2013 and Neri et al., 2006 also reported that merely observing the individual actions of one's partner helped to shape predictions about that partner's actions with high temporal and spatial precision.

An important caution is that the behavioral evidence described above does not offer a strong answer to the question of *how* self and other predictive processes are organized. Determining this will likely only be possible in the future by implementing PJAM's theoretical ideas in a fully functional computational model. Such a transition, from written theory to dynamic modeling, will make clear in a concrete way where PJAM's description needs greater specification.

Limitations and future work

PJAM offers one answer to the question posed by Sebanz and Knoblich (2009). These authors concluded that "One main challenge for future work seems to be to understand how lower-level processes like action simulation and higher-level processes like verbal communication and mental state attribution work in concert, and under which circumstances they can overrule each other." PJAM offers a solution by explicitly linking processes occurring at different levels (e.g., sensory, motor, and symbolic).

Some findings that are not readily explained by PJAM concern the interaction of motor mechanisms with contextual and personal factors. For example, a partner's tendency to represent the other partner's task can be modulated by an ever-growing list of variables, including mood (Kuhbandner et al., 2010), joint goals (de Bruijn

et al., 2008; Iani et al., 2011), self-construal (Colzato, de Bruijn, et al., 2012), religious belief (Colzato, Zech, et al., 2012), social power structure (Hogeveen, Inzlicht, & Obhi, 2014), personality traits (Hogeveen & Obhi, 2013), primed social concepts (Obhi & Sebanz, 2011), long-term social bonds (Kang et al., 2010), and artificially created social affiliations (Wenke et al., 2011). Further research is necessary to determine whether these factors operate at the sensory routing, action-planning, and/or goal representation layer. In Box 1, we list the questions for future research that were identified earlier in this review, bringing them together in a convenient single location.

Box 1. Questions for future research.

1. How are physical effort (the total energy required to carry out a task) and coordination effort (the mental effort involved in predicting a partner's actions and adjusting one's own actions accordingly) weighted to achieve optimal social cooperation?
2. How is the high-level experience of agency within a joint-action related to low-level discrimination of self and other action outcomes?
3. Are internal motor models of self and other construed independently or contingently from one another?
4. If self and other models are construed independently, do we then need to have internal joint-models in order to produce predictions about the synergistic effects of self and other actions?
5. What are the mechanisms underlying the influence of context-specific and person-specific factors on the execution of joint-actions?
6. Can perceptual predictions about partners be achieved without motor simulation?
7. What is the limit number of partners that can be internally represented by the motor system?
8. What are the processes that regulate the gain of prediction errors pertaining to self and other within a joint-action?

In closing, we propose that PJAM is a useful organizing framework for the study of joint-action. The framework elaborates joint-action through hierarchical predictive principles, which sharpens the theoretical rigor of previous accounts (Vesper et al., 2010; Wolpert et al., 2003). As such, it offers an overarching organizational framework for the diverse findings in joint-action literature. In making this comparison, several areas of research were identified that are needed to provide finer-grained detail for the framework. We also believe there is enough promise in this approach to encourage PJAM as a stepping-stone to the construction of fully computational models of joint-action.

Acknowledgments This work was supported by a Ph.D. scholarship to author A. Pesquita from the Portuguese Fundação para a Ciência e Tecnologia (SFRH/BD/76087/2011), by a post-doctoral fellowship to R. Whitwell from the Natural Science and Engineering Council of Canada, and by a Discovery Grant to J. T. Enns from the Natural Science and Engineering Council of Canada. The authors are grateful to Ulysses Bernardet, Rebecca Todd, Alan Kingston, Mark Schaller, Peter Graf, and Karon Maclean for comments and discussions that helped shape the manuscript.

References

- Atmaca, S., Sebanz, N., & Knoblich, G. (2011). The joint flanker effect: Sharing tasks with real and imagined co-actors. *Experimental Brain Research*, 211(3–4), 371–85. <https://doi.org/10.1007/s00221-011-2709-9>
- Atmaca, S., Sebanz, N., Prinz, W., & Knoblich, G. (2008). Action co-representation: The joint SNARC effect. *Social Neuroscience*, 3(3–4), 410–20. <https://doi.org/10.1080/17470910801900908>
- Baron, R. A. (1987). Interviewer's moods and reactions to job applicants: The Influence of affective states on applied social judgments. *Journal of Applied Social Psychology*, 17, 911–926. <https://doi.org/10.1111/j.1559-1816.1987.tb00298.x>
- Becchio, C., Sartori, L., & Castiello, U. (2010). Toward you: The social side of actions. *Current Directions in Psychological Science*, 19(3), 183–188. <https://doi.org/10.1177/0963721410370131>
- Blakemore, S. J., & Frith, C. (2003). Self-awareness and action. *Current Opinion in Neurobiology*. [https://doi.org/10.1016/S0959-4388\(03\)00043-6](https://doi.org/10.1016/S0959-4388(03)00043-6)
- Blakemore, S. J., Frith, C. D., & Wolpert, D. M. (1999). Spatio-temporal prediction modulates the perception of self-produced stimuli. *Journal of Cognitive Neuroscience*, 11(5), 551–559. <https://doi.org/10.1162/089892999563607>
- Blakemore, S. J., Goodbody, S. J., & Wolpert, D. M. (1998). Predicting the consequences of our own actions: The role of sensorimotor context estimation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 18, 7511–7518. Not available
- Blakemore, S. J., Wolpert, D., & Frith, C. (2000). Why can't you tickle yourself? *Neuroreport*, 11(11), 11–16. <https://doi.org/10.1586/14737175.7.10.1337>
- Bratman, M. E. (1992). Shared cooperative activity. *Philosophical Review*, 101(2), 327–341.
- Brown, H., Adams, R. A., Parees, I., Edwards, M., & Friston, K. (2013). Active inference, sensory attenuation and illusions. *Cognitive Processing*, 14(4), 411–427. <https://doi.org/10.1007/s10339-013-0571-3>
- Carp, J., Halenar, M. J., Quandt, L. C., Sklar, A., & Compton, R. J. (2009). Perceived similarity and neural mirroring: Evidence from vicarious error processing. *Social Neuroscience*, 4(1), 85–96. <https://doi.org/10.1080/17470910802083167>
- Chaminade, T., Marchant, J. L., Kilner, J., & Frith, C. D. (2012). An fMRI study of joint action-varying levels of cooperation correlates with activity in control networks. *Front Hum Neurosci*, 6(June), 179. <https://doi.org/10.3389/fnhum.2012.00179>
- Clark, A. (2013). Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181–253. <https://doi.org/10.1017/S0140525X12000477>
- Clark, H. H. (1996). *Using language*. Cambridge, England: Cambridge University Press.
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. *Perspectives on Socially Shared Cognition*. <https://doi.org/10.1037/10096-006>
- Colzato, L. S., de Bruijn, E. R. A., & Hommel, B. (2012). Up to “me” or up to “us”? The impact of self-construal priming on cognitive self-other integration. *Frontiers in Psychology*, 3(September), 341. <https://doi.org/10.3389/fpsyg.2012.00341>
- Colzato, L. S., Zech, H., Hommel, B., Verdonschot, R., van den Wildenberg, W. P. M., & Hsieh, S. (2012). Loving-kindness brings loving-kindness: The impact of Buddhism on cognitive self-other integration. *Psychonomic Bulletin & Review*, 19(3), 541–5. <https://doi.org/10.3758/s13423-012-0241-y>
- de Bruijn, E. R., Miedl, S. F., & Bekkering, H. (2008). Fast responders have blinders on: ERP correlates of response inhibition in competition. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, 44(5), 580–6. <https://doi.org/10.1016/j.cortex.2007.09.003>
- Doerrfeld, A., Sebanz, N., & Shiffrar, M. (2012). Expecting to lift a box together makes the load look lighter. *Psychological Research*, 76(4), 467–75. <https://doi.org/10.1007/s00426-011-0398-4>
- Dolk, T., Hommel, B., Prinz, W., & Liepelt, R. (2013). The (not so) social Simon effect: A referential coding account. *Journal of Experimental Psychology: Human Perception and Performance*, 39(5), 1248–60. <https://doi.org/10.1037/a0031031>
- Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, 424(6950), 769–71. <https://doi.org/10.1038/nature01861>
- Flanagan, R. R., Vetter, P., Johansson, R. S., & Wolpert, D. M. (2003). Prediction precedes control in motor learning. *Current Biology*, 13, 146–150. [https://doi.org/10.1016/S0960-9822\(03\)00007-1](https://doi.org/10.1016/S0960-9822(03)00007-1)
- Forgas, J. P. (1998). On feeling good and getting your way: Mood effects on negotiator cognition and bargaining strategies. *Journal of Personality and Social Psychology*, 74, 565–577. <https://doi.org/10.1037/0022-3514.74.3.565>
- Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review*, 5(3), 490–495. <https://doi.org/10.3758/BF03208827>
- Friston, K. (2003). Learning and inference in the brain. *Neural Networks*, 16(9), 1325–1352. <https://doi.org/10.1016/j.neunet.2003.06.005>
- Friston, K. (2008). Hierarchical models in the brain. *PLoS Computational Biology*, 4(11). <https://doi.org/10.1371/journal.pcbi.1000211>
- Friston, K., & Frith, C. (2015a). A duet for one. *Consciousness and Cognition*, 36, 390–405. <https://doi.org/10.1016/j.concog.2014.12.003>
- Friston, K., Mattout, J., & Kilner, J. (2011). Action understanding and active inference. *Biological Cybernetics*, 104(1–2), 137–60. <https://doi.org/10.1007/s00422-011-0424-z>
- Friston, K. J., & Frith, C. D. (2015b). Active inference, Communication and hermeneutics. *Cortex*, 68(Kelso 2012), 129–143. <https://doi.org/10.1016/j.cortex.2015.03.025>
- Frith, C. D., Blakemore, S. J., & Wolpert, D. M. (2000). Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. In *Brain Research Reviews* (Vol. 31, pp. 357–363). [https://doi.org/10.1016/S0165-0173\(99\)00052-1](https://doi.org/10.1016/S0165-0173(99)00052-1)
- Goebel, W., & Palmer, C. (2009). Synchronization of timing and motion among performing musicians. *Music Perception*, 427–438. <https://doi.org/10.1525/mp.2009.26.5.427>
- Graf, M., Reitzner, B., Corves, C., Casile, A., Giese, M., & Prinz, W. (2007). Predicting point-light actions in real-time. *NeuroImage, 36 Suppl 2*, T22-32. <https://doi.org/10.1016/j.neuroimage.2007.03.017>
- Graziano, M. S. A. (2013). *Consciousness and the social brain*. New York: Oxford University Press.
- Gredebäck, G., & Falck-Ytter, T. (2015). Eye movements during action observation. *Perspectives in Cognitive Science*, 10(5), 591–598. <https://doi.org/10.1177/1745691615589103>
- Häberle, A., Schütz-Bosbach, S., Laboisière, R., & Prinz, W. (2008). Ideomotor action in cooperative and competitive settings. *Social Neuroscience*, 3(1), 26–36. <https://doi.org/10.1080/17470910701482205>
- Haruno, M., Wolpert, D. M., & Kawato, M. (2003). Hierarchical MOSAIC for movement generation. *International Symposium on Limbic and Association Cortical Systems*, 1250, 575–590. [https://doi.org/10.1016/S0531-5131\(03\)00190-0](https://doi.org/10.1016/S0531-5131(03)00190-0)
- Hogeveen, J., Inzlicht, M., & Obhi, S. S. (2014). Power changes how the brain responds to others. *Journal of Experimental Psychology: General*, 143(2), 755–62. <https://doi.org/10.1037/a0033477>
- Hogeveen, J., & Obhi, S. S. (2013). Automatic imitation is automatic, but less so for narcissists. *Experimental Brain Research*, 224, 613–621. <https://doi.org/10.1007/s00221-012-3339-6>
- Holländer, A., Jung, C., & Prinz, W. (2011). Covert motor activity on NoGo trials in a task sharing paradigm: Evidence from the

- lateralized readiness potential. *Experimental Brain Research*, 211(3–4), 345–56. <https://doi.org/10.1007/s00221-011-2688-x>
- Iani, C., Anelli, F., Nicoletti, R., Arcuri, L., & Rubichi, S. (2011). The role of group membership on the modulation of joint action. *Experimental Brain Research*, 211(3–4), 439–45. <https://doi.org/10.1007/s00221-011-2651-x>
- Kang, S. K., Hirsh, J. B., & Chasteen, A. L. (2010). Your mistakes are mine: Self-other overlap predicts neural response to observed errors. *Journal of Experimental Social Psychology*, 46(1), 229–232. <https://doi.org/10.1016/j.jesp.2009.09.012>
- Keller, P. E. (2008). Joint action in music performance. In *Enacting intersubjectivity: A cognitive and social perspective on the study of interactions* (pp. 205–221).
- Keller, P. E., Knoblich, G., & Repp, B. H. (2007). Pianists duet better when they play with themselves: On the possible role of action simulation in synchronization. *Consciousness and Cognition*, 16(1), 102–111. <https://doi.org/10.1016/j.concog.2005.12.004>
- Keller, P. E., Novembre, G., & Loehr, J. D. (2016). Musical ensemble performance: Representing self, other and joint action outcomes. In S. S. Obhi & E. S. Cross (Eds.), *Shared representations: Sensorimotor foundations of social life* (p. 280). Cambridge, England: Cambridge University Press.
- Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological research on joint action: Theory and data. In WDK2003 (Ed.), *The psychology of learning and motivation* (Vol. 54, pp. 59–101). Burlington: Academic Press. <https://doi.org/10.1016/B978-0-12-385527-5.00003-6>
- Knoblich, G., & Jordan, J. S. (2003). Action coordination in groups and individuals: Learning anticipatory control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(5), 1006–16. <https://doi.org/10.1037/0278-7393.29.5.1006>
- Knoblich, G., & Sebanz, N. (2006). The social nature of perception and action. *Current Directions in Psychological Science*, 15(3), 99–104. <https://doi.org/10.1111/j.0963-7214.2006.00415.x>
- Koban, L., Pourtois, G., Vocat, R., & Vuilleumier, P. (2010). When your errors make me lose or win: event-related potentials to observed errors of cooperators and competitors. *Social Neuroscience*, 5(4), 360–74. <https://doi.org/10.1080/17470911003651547>
- Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: cognitive basis for stimulus-response compatibility—A model and taxonomy. *Psychological Review*, 97, 253–270. <https://doi.org/10.1037/0033-295X.97.2.253>
- Kourtis, D., Sebanz, N., & Knoblich, G. (2010). Favouritism in the motor system: social interaction modulates action simulation. *Biology Letters*, 6(6), 758–61. <https://doi.org/10.1098/rsbl.2010.0478>
- Kubbandner, C., Pekrun, R., & Maier, M. a. (2010). The role of positive and negative effect in the “mirroring” of other persons’ actions. *Cognition & Emotion*, 24(7), 1182–1190. <https://doi.org/10.1080/02699930903119196>
- Langton, S. R., & Bruce, V. (2000). You must see the point: Automatic processing of cues to the direction of social attention. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 747–757. <https://doi.org/10.1037/0096-1523.26.2.747>
- Liepelt, R., & Prinz, W. (2011). How two share two tasks: Evidence of a social psychological refractory period effect. *Experimental Brain Research*, 211(3–4), 387–96. <https://doi.org/10.1007/s00221-011-2703-2>
- Loehr, J. D., Kourtis, D., Vesper, C., Sebanz, N., & Knoblich, G. (2013). Monitoring individual and joint action outcomes in duet music performance. *Journal of Cognitive Neuroscience*, 25(7), 1049–61. https://doi.org/10.1162/jocn_a_00388
- Loehr, J. D., & Vesper, C. (2016). The sound of you and me: Novices represent shared goals in joint action. *Quarterly Journal of Experimental Psychology*, 69(3), 535–547. <https://doi.org/10.1080/17470218.2015.1061029>
- Manera, V., Schouten, B., Verfaillie, K., & Becchio, C. (2013). Time will show: Real time predictions during interpersonal action perception. *PLoS One*, 8(1), e54949. <https://doi.org/10.1371/journal.pone.0054949>
- Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural Networks*. [https://doi.org/10.1016/S0893-6080\(96\)00035-4](https://doi.org/10.1016/S0893-6080(96)00035-4)
- Neri, P., Luu, J. Y., & Levi, D. M. (2006). Meaningful interactions can enhance visual discrimination of human agents. *Nature Neuroscience*, 9(9), 1186–92. <https://doi.org/10.1038/nn1759>
- Novembre, G., Ticini, L. F., Schütz-Bosbach, S., & Keller, P. E. (2012). Distinguishing self and other in joint action. Evidence from a musical paradigm. *Cerebral Cortex (New York, N.Y.: 1991)*, 22(12), 2894–903. <https://doi.org/10.1093/cercor/bhr364>
- Numminen, J., Salmelin, R., & Hari, R. (1999). Subject’s own speech reduces reactivity of the human auditory cortex. *Neuroscience Letters*, 265(2), 119–122. [https://doi.org/10.1016/S0304-3940\(99\)00218-9](https://doi.org/10.1016/S0304-3940(99)00218-9)
- Obhi, S. S. (2012). The troublesome distinction between self-generated and externally triggered action: A commentary on Schüür and Haggard. *Consciousness and Cognition*, 21(1), 587–8. <https://doi.org/10.1016/j.concog.2011.09.014>
- Obhi, S. S., & Hall, P. (2011). Sense of agency and intentional binding in joint action. *Experimental Brain Research*, 211(3–4), 655–62. <https://doi.org/10.1007/s00221-011-2675-2>
- Obhi, S. S., & Sebanz, N. (2011). Moving together: Toward understanding the mechanisms of joint action. *Experimental Brain Research*, 211(3–4), 329–36. <https://doi.org/10.1007/s00221-011-2721-0>
- Pacherie, E. (2012). The phenomenology of joint action: Self-agency vs. joint-agency. *Joint Attention: New Developments*, 343–389. <https://doi.org/10.1007/s13164-011-0052-5>
- Parkinson, J., Springer, A., & Prinz, W. (2012). Before, during and after you disappear: Aspects of timing and dynamic updating of the real-time action simulation of human motions. *Psychological Research*, 76, 421–433. <https://doi.org/10.1007/s00426-012-0422-3>
- Pecenka, N., & Keller, P. E. (2011). The role of temporal prediction abilities in interpersonal sensorimotor synchronization. *Experimental Brain Research*, 211(3–4), 505–15. <https://doi.org/10.1007/s00221-011-2616-0>
- Pickering, M. J., & Clark, A. (2014). Getting ahead: Forward models and their place in cognitive architecture. *Trends in Cognitive Sciences*. Elsevier Ltd. <https://doi.org/10.1016/j.tics.2014.05.006>
- Radke, S., de Lange, F. P., Ullsperger, M., & de Bruijn, E. R. A. (2011). Mistakes that affect others: An fMRI study on processing of own errors in a social context. *Experimental Brain Research*, 211(3–4), 405–13. <https://doi.org/10.1007/s00221-011-2677-0>
- Ramenzoni, V. C., Sebanz, N., & Knoblich, G. (2014). Scaling up perception-action links: Evidence from synchronization with individual and joint action. *Journal of Experimental Psychology: Human Perception and Performance*, 40(4), 1551–65. <https://doi.org/10.1037/a0036925>
- Ramnani, N., & Miall, R. C. (2004). A system in the human brain for predicting the actions of others. *Nature Neuroscience*, 7(1), 85–90. <https://doi.org/10.1038/nn1168>
- Rizzolatti, G., Fogassi, L., & Gallese, V. (2002). Motor and cognitive functions of the ventral premotor cortex. *Current Opinion in Neurobiology*, 12, 149–154.
- Sartori, L., Becchio, C., & Castiello, U. (2011). Cues to intention: The role of movement information. *Cognition*, 119(2), 242–52. <https://doi.org/10.1016/j.cognition.2011.01.014>
- Sato, A. (2008). Action observation modulates auditory perception of the consequence of others’ actions. *Consciousness and Cognition*, 17(4), 1219–27. <https://doi.org/10.1016/j.concog.2008.01.003>
- Savelsbergh, G. J. P., Williams, A. M., Van der Kamp, J., & Ward, P. (2002). Visual search, anticipation and expertise in soccer

- goalkeepers. *Journal of Sports Sciences*, 20, 279–287. <https://doi.org/10.1080/026404102317284826>
- Schüür, F., & Haggard, P. (2012). On capturing the essence of self-generated action: A reply to Obhi (2012). *Consciousness and Cognition*, 21(2), 1070–1. <https://doi.org/10.1016/j.concog.2012.03.003>
- Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science*, 1(2), 353–367. <https://doi.org/10.1111/j.1756-8765.2009.01024.x>
- Sebanz, N., & Shiffrar, M. (2007). Bodily bonds: Effects of social context on ideomotor movements. In Y. Rossetti, M. Kawato, & P. Haggard (Eds.), *Sensorimotor foundations of higher cognition (attention and performance, XXII)*. Oxford, UK: Oxford University Press.
- Sebanz, N., & Shiffrar, M. (2009). Detecting deception in a bluffing body: The role of expertise. *Psychonomic Bulletin & Review*, 16(1), 170–5. <https://doi.org/10.3758/PBR.16.1.170>
- Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, 10(2), 70–6. <https://doi.org/10.1016/j.tics.2005.12.009>
- Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: Just like one's own? *Cognition*, 88(3), B11–21. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12804818>
- Sebanz, N., Knoblich, G., & Prinz, W. (2005). How two share a task: Corepresenting stimulus-response mappings. *Journal of Experimental Psychology: Human Perception and Performance*, 31(6), 1234–46. <https://doi.org/10.1037/0096-1523.31.6.1234>
- Sebanz, N., Knoblich, G., Prinz, W., & Wascher, E. (2006). Twin peaks: An ERP study of action planning and control in co-acting individuals. *Journal of Cognitive Neuroscience*, 18(5), 859–70. <https://doi.org/10.1162/jocn.2006.18.5.859>
- Semin, R., & Cacioppo, J. T. (2006). Synchronization, coordination, and co-regulation. In *Grounding social cognition* (pp. 119–128).
- Simon, J. R. (1969). Reactions toward the source of stimulation. *Journal of Experimental Psychology*, 81(1), 174–176. <https://doi.org/10.1037/h0027448>
- Simon, J. R., & Craft, J. L. (1970). Effects of an irrelevant auditory stimulus on visual choice reaction time. *Journal of Experimental Psychology*, 86, 272–274. <https://doi.org/10.1037/h0029961>
- Sparenberg, P., Springer, A., & Prinz, W. (2012). Predicting others' actions: Evidence for a constant time delay in action simulation. *Psychological Research*, 76(1), 41–9. <https://doi.org/10.1007/s00426-011-0321-z>
- Springer, A., Hamilton, A. F. D. C., & Cross, E. S. (2012). Simulating and predicting others' actions. *Psychological Research*, 76(4), 383–7. <https://doi.org/10.1007/s00426-012-0443-y>
- Stix, G. (2014). The “It” factor. *Scientific American*, 311(3), 72–79. <https://doi.org/10.1038/scientificamerican0914-72>
- Todorov, E. (2004). Optimality principles in sensorimotor control. *Nature Neuroscience*, 7(9), 907–15. <https://doi.org/10.1038/nn1309>
- Tomasello, M. (2009). *Why we cooperate*. *Human Resource Management* (Vol. 49). <https://doi.org/10.1002/hrm.20395>
- Tsai, C.-C., Kuo, W.-J., Jing, J.-T., Hung, D. L., & Tzeng, O. J.-L. (2006). A common coding framework in self-other interaction: Evidence from joint action task. *Experimental Brain Research*, 175(2), 353–62. <https://doi.org/10.1007/s00221-006-0557-9>
- van der Wel, R. P. R. D., Sebanz, N., & Knoblich, G. (2012). The sense of agency during skill learning in individuals and dyads. *Consciousness and Cognition*, 21(3), 1267–79. <https://doi.org/10.1016/j.concog.2012.04.001>
- van Schie, H. T., Mars, R. B., Coles, M. G. H., & Bekkering, H. (2004). Modulation of activity in medial frontal and motor cortices during error observation. *Nature Neuroscience*, 7(5), 549–54. <https://doi.org/10.1038/nn1239>
- Vesper, C., Butterfill, S., Knoblich, G., & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks: The Official Journal of the International Neural Network Society*, 23(8–9), 998–1003. <https://doi.org/10.1016/j.neunet.2010.06.002>
- Vesper, C., van der Wel, R. P. R. D., Knoblich, G., & Sebanz, N. (2011). Making oneself predictable: Reduced temporal variability facilitates joint action coordination. *Experimental Brain Research*, 211(3–4), 517–30. <https://doi.org/10.1007/s00221-011-2706-z>
- Vesper, C., van der Wel, R. P. R. D., Knoblich, G., & Sebanz, N. (2013). Are you ready to jump? Predictive mechanisms in interpersonal coordination. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1), 48–61. <https://doi.org/10.1037/a0028066>
- Wegner, D. M., Sparrow, B., & Winerman, L. (2004). Vicarious agency: Experiencing control over the movements of others. *Journal of Personality and Social Psychology*, 86(6), 838–48. <https://doi.org/10.1037/0022-3514.86.6.838>
- Wenke, D., Atmaca, S., Holländer, A., Liepelt, R., Baess, P., & Prinz, W. (2011). What is shared in joint action? Issues of co-representation, response conflict, and agent identification. *Review of Philosophy and Psychology*, 2(2), 147–172. <https://doi.org/10.1007/s13164-011-0057-0>
- Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, 131(3), 460–73. <https://doi.org/10.1037/0033-2909.131.3.460>
- Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in Cognitive Sciences*, 1(6), 209–16. [https://doi.org/10.1016/S1364-6613\(97\)01070-X](https://doi.org/10.1016/S1364-6613(97)01070-X)
- Wolpert, D. M., Doya, K., & Kawato, M. (2003). 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. <https://doi.org/10.1098/rstb.2002.1238>
- Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology*, 11(18), R729–R732. [https://doi.org/10.1016/S0960-9822\(01\)00432-8](https://doi.org/10.1016/S0960-9822(01)00432-8)
- Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, 269, 1880–1882. <https://doi.org/10.1126/science.7569931>
- Wolpert, D. M., & Miall, R. C. (1996). Forward models for physiological motor control. *Neural Networks: The Official Journal of the International Neural Network Society*, 9(8), 1265–1279. [https://doi.org/10.1016/S0893-6080\(96\)00035-4](https://doi.org/10.1016/S0893-6080(96)00035-4)
- Wolpert, D. M., Miall, R. C., Britain, G., Senior, W., & Trust, W. (1996). Forward models for physiological motor control. *Neural Networks: The Official Journal of the International Neural Network Society*, 9(8), 1265–1279. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12662535>