A Predictive Processing Framework for Joint Action and Communication

Giovanni Pezzulo[1,*], Günther Knoblich[2], Domenico Maisto[1], Francesco Donnarumma[1], Elisabeth Pacherie[3], Uri Hasson[4]

1) Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy
2) Department of Cognitive Science, Central European University, Vienna, Austria
3) Institut Jean-Nicod, EHESS, CNRS, École Normale Supérieure, PSL University, Paris, France
4) Department of Psychology, Princeton University, Princeton, NJ, USA

* Corresponding author: giovanni.pezzulo@istc.cnr.it

Abstract

Humans act together to achieve feats they could never achieve alone and communicate to ensure alignment of meaning and understanding across different individuals. Explaining the unique human joint action and communication abilities poses an enormous challenge because it requires a systematic account of how people go beyond their own individual perceptions, thoughts, and needs to achieve joint outcomes and align their understanding. Here, we advance a new theoretical framework for explaining joint action and communication. It builds upon influential predictive processing architectures, extending them from individual cognition to multiagent, interactive settings. We assume that joint action and communication involve using and updating agent-neutral models that enable co-agents to predict collective outcomes of interactions regardless of who achieved them. This contrasts with previous frameworks postulating that agent-specific models predict action outcomes for self and others. We discuss three key claims derived from our framework: 1) Co-agents use agent-neutral predictive frameworks during joint action; 2) Co-agents update agent-neutral models interactively by shaping others' predictions through verbal and non-verbal communication; and 3) Agent-neutral models enable dynamic role allocation during joint action. We highlight how these three claims stem from our proposal, what evidence currently favors or disfavors them, and what novel experiments could be conducted to test them further. Our agent-neutral predictive processing framework will provide a new perspective for understanding the individual basis of human sociality, which closely links theories of joint action and communication to principles of computational neuroscience.

Keywords: predictive processing; joint action; coordination; communication; agent-neutral models

Introduction

Humans act together to achieve feats that go much beyond individual abilities. The human propensity to engage in joint action is a defining feature of our human sociality [1–4]. It is the key ingredient to achieving social organization [5,6], verbal communication [7], and cultural transmission [8,9]. Humans' ability to engage in instrumental and communicative joint actions vastly expands the impact that human action can have on the world and the goals and beliefs that can be shared with others. Joint action gives rise to a rich set of unique subjective experiences, including a sense of joint agency over action outcomes that are beyond individual control [10,11] and a sense of commitment to go through with joint actions [12,13]. Furthermore, joint action is at the core of humans' ability to have conversations in natural language, a highly flexible and powerful means for aligning meaning and understanding across different individuals [14–17].

Understanding joint action and communication in the wide variety of forms it manifests poses an enormous challenge in Cognitive Science, Neuroscience, and the Social Sciences because it requires an explanation of how individuals go beyond their perceptions, thoughts, and needs to achieve and maintain alignment with others to work toward a joint outcome or understanding. Previous explanations of joint action and communication have highlighted particular types or levels of alignment. Prominent theories in the philosophy of action have highlighted the role of shared or collective intentions [18,19] and joint commitment [20]. Theorizing in pragmatics and cognitive linguistics has highlighted the role of language use [7] and ostension in communication [21]. Economists argue that joint payoff can overrule individual payoff, following the logic of team reasoning [22,23]. Work in cognitive neuroscience has identified temporal entrainment and neural coupling [15,24,25], and mirroring [26,27] as crucial mechanisms for explaining how tight spatiotemporal coordination is achieved in joint action. Recently, there has been a move towards using predictive processing models to explain linguistic alignment [28], virtual bargaining [29], action understanding [30], joint action coordination [31], sensorimotor communication [32], sense of agency [33–35], and sense of commitment [13]. A shared assumption between these models is that individuals create predictive models for their own actions and beliefs, as well as those of their partners.

Here, we advance a new unified computational framework for joint action and communication that extends the highly influential predictive processing architectures from individual cognition [36–38] to multiagent, interactive settings. Our integrated predictive processing framework for joint action and communication builds on computational principles of predictive processing that have led to fundamental advances in understanding the dynamics of individual perception and action and the nature of language processing in the human mind and brain.

The fundamental idea distinguishing our new proposal from previous attempts is that of a transition from an agent-centered perspective to an agent-neutral perspective. Our new *agent-neutral predictive framework* assumes that during joint action, individuals transition from using agent-specific models that predict action outcomes for self and others to a shared, agent-neutral predictive model that predicts the collective consequences of their aggregate actions, regardless of each individual's contribution. Figure 1 shows schematically the key differences between these two alternative views of predictive modeling in social interaction and joint action, based upon agent-specific versus agent-neutral predictive models.

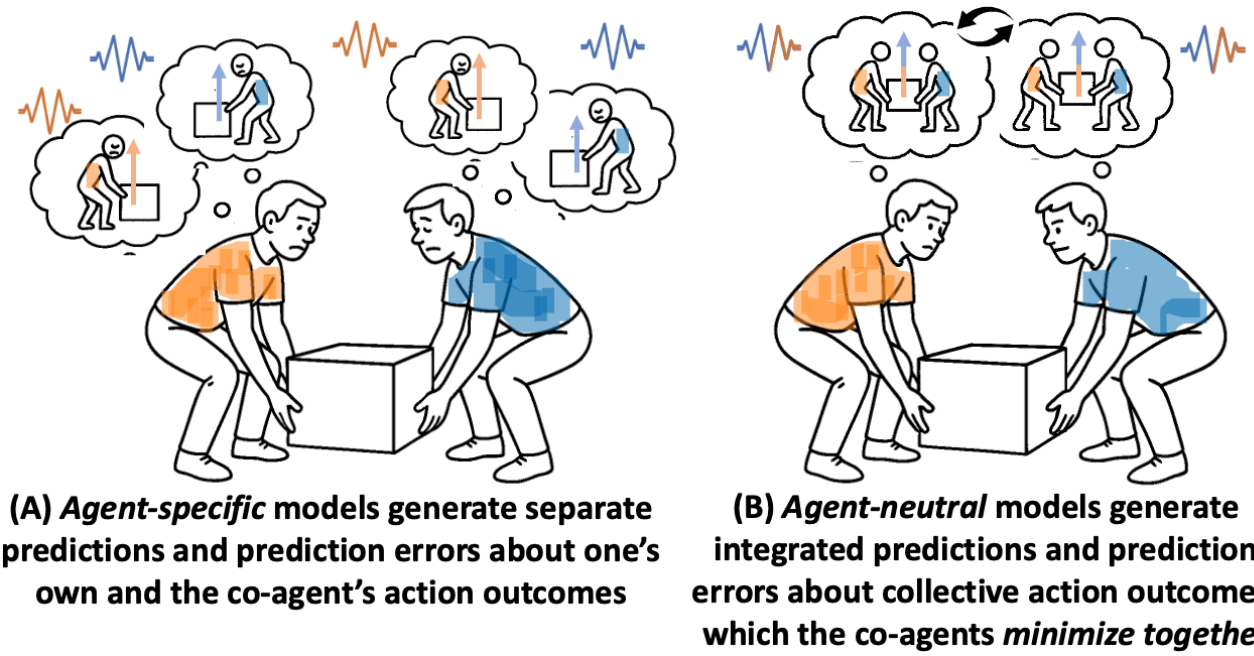

Figure 1. *Schematic illustration of two accounts of predictive modeling in social interaction and joint action: agent-specific versus agent-neutral models. (A) The traditional view holds that individuals use agent-specific internal models to separately control and predict their own actions and those of others* [30,39–42]. *Joint action, such as lifting an object together, can be divided into two phases (often interleaved). In phase one (action observation), co-agent-1 reuses their motor-control models to understand and predict co-agent-2's actions (e.g., inferring their intention to lift and predicting the height reached). In phase two (joint planning), co-agent-1 uses their internal model to plan a complementary action—for instance, adding the missing force to reach the target height. A natural extension for strategic interaction involves intention inference (mindreading): predicting the impact of one's actions on the partner's future beliefs and behavior, possibly recursively (e.g., what co-agent-2 will believe and do if I act this way; what they believe I will believe, etc.)* [43]. *The simulated EEG waveforms illustrate agent-specific predictions and errors, color-coded by co-agent identity. (B) In contrast, the agent-neutral approach uses a model that infers task-relevant variables jointly controlled by both agents, such as total lifting force. It predicts collective outcomes (e.g., object height based on combined forces), tracks discrepancies between predicted and desired outcomes, and supports agent-neutral plans—plans that achieve the joint goal regardless of who executes them. Examples of this collective-control view (not always framed in predictive-processing terms) include* [44–48]. *In the figure, simulated EEG waveforms mark agent-neutral predictions and errors, while reciprocal arrows depict how agent-neutral models align and become shared over time.*

In the following three sections, we introduce three key claims stemming from our new agent-neutral perspective. See Figure 2 for a schematic illustration of the three claims, in the hypothetical scenario of two co-agents building Lego towers together. For each claim, we explain the rationale within predictive processing, discuss the supporting empirical evidence, and suggest possible future experiments that test the claims.

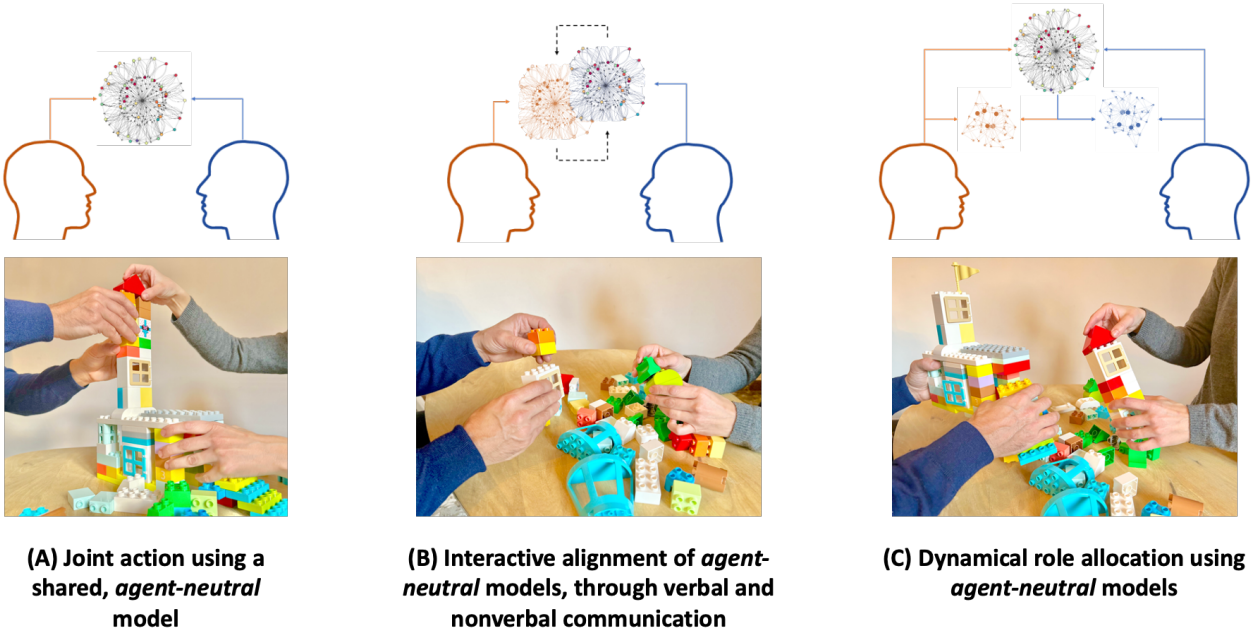

Figure 2. A schematic illustration of three claims from the agent-neutral perspective. *(A) Joint action: When two (or more) agents act together, they use an agent-neutral predictive model—and a shared understanding of the task—to coordinate. In the top panel, the model is depicted as a "graph" of sequential actions needed to build a tower, independent of who performs each step. Nodes are colored bricks, edges are assembly steps. The shared graph guides both prediction and action. The bottom panel shows that coordination also requires mutual support: co-agents generate similar predictions about continuing the tower, and either can fulfill them by adding pieces—thus reducing prediction errors for both. (B) Interactive alignment: During interaction, co-agents align and update their agent-neutral models through sensorimotor and verbal communication, reducing uncertainty about "which tower we should build." The top panel shows schematic alignment of two graphs; the bottom shows co-actors deciding what to construct when the outcome is still open. Alignment may involve temporarily violating expectations to compel updating, creating transient surprise that promotes joint success and reduces future uncertainty. (B) Dynamic role allocation: Agent-neutral modeling supports dynamic role distribution and goal assignment. The top panel shows subgraphs corresponding to different construction parts; the bottom shows one agent building the bottom floor, the other the top, later combining them.*

First claim: Co-agents use a shared agent-neutral predictive model to predict and optimize collective outcomes and joint actions.

Many joint actions like carrying a sofa up the stairs, playing soccer, dancing a tango, playing a duet, or improvising together require precise and dynamic spatiotemporal coordination between the individuals performing the joint action [49,50]. The traditional approach to explain joint action coordination uses internal predictive models to predict one's actions and their reuse to predict other agents' actions and infer their intentions [30]. Instead, in an agent-neutral perspective, co-agents depend on a shared, agent-neutral understanding of the world to coordinate their actions. Crucially, the traditional approach of using separate models for one's own and other agents' actions is replaced by a single agent-neutral model that predicts the combined results of joint actions, regardless of who carries them out. In predictive processing terms, using agent-neutral models, two (or more) co-agents can collectively minimize the same prediction error about collective outcomes rather than combining the predictions and prediction errors of two models, one for my actions and one for others' actions, as assumed in the traditional view. For example, when lifting a table together, an agent-neutral model would allow for the prediction of the movements of the table that result from the combination of the forces applied by both agents. Using agent-neutral models improves joint action by alleviating cognitive demands because it does not require making and integrating multiple agent-specific predictions to infer how the table will move. It allows each agent to make flexible online adjustments to the joint plan, e.g., one agent can compensate for the other agent's lack of force. It also scales up when more co-agents become involved in the joint action, which would be more problematic if forming and simultaneously using multiple agent-specific models, one for each co-agent, were required. Hence, the assumption of agent-neutral models explains the surprising efficiency of human joint action in terms of a more efficacious prediction error minimization.

A fundamental assumption of predictive processing and active inference is that the brain forms an internal (generative) model of the environment and its regularities and uses it to infer and predict the state of the world and the best course of action to achieve goals [36–38]. While most computational models of social interaction using predictive processing (or related approaches) assume separate, agent-specific models for oneself and others, some have implemented aspects of agent-neutrality. For example, in the "duet for one" model [44], two co-agents (depicted as "birds") monitor the same collective, agent-neutral variable: both expect to hear a preferred melody, regardless of which bird sings. Each bird has an identical generative model enabling both production and perception of the melody. The key insight is that if one bird sings, the other's prediction is fulfilled without acting. A simulation shows that in an initial phase, when the birds cannot hear each other, their expectations evolve independently. Once auditory contact is established, expectations rapidly align. This sharing of posterior expectations ("generalized synchrony") marks interactive alignment, which only emerges with mutual auditory access and breaks down if information flow is interrupted. Another model that incorporates aspects of agent-neutrality is called "interactive inference" [45]. Here, two co-agents navigate a maze from opposite starting points to jointly press either a red or blue button. Their generative models include both agent-specific and agent-neutral components. The agent-neutral part maintains beliefs about the joint goal (e.g., both pressing red or blue), updated by observing both agents' actions. In a "leaderless" scenario, with no strong initial preferences, the agent-neutral model supports aligned behavior and shared goal formation. In a "leader–follower" scenario, one agent (leader) knows the joint goal, while the other (follower) does not. The leader uses non-verbal, sensorimotor communication [32] to help the follower infer the goal—even if this requires a costlier plan, such as taking a longer trajectory that clearly indicates the correct button. This follows from policy selection depending on expected free energy, which balances two objectives: choosing actions that realize preferred goals and making "social epistemic actions" that reduce uncertainty about the joint goal. Other related modeling studies propose that during interactive inference, a person's lower-level neural mechanisms for sensorimotor transformation and spatial representation could be "recalibrated" in a social manner to incorporate new information relative to the other agent(s) - analogous to the recalibration of visual receptive fields in response to repeated tool use [51]. Through dynamic alignment, this mechanism could facilitate the creation of *agent-neutral* models. Yet other computational approaches aim to understand the collective behavior of multiple interactive co-agents by inferring causal models of how co-agents relate and their collective plans [52], the sub-tasks others are working on [53], or the (average) collective decisions of multiple co-agents [54]. Other computational models represent agent-neutral models as an "imagined we," inferring the plans a "we" agent would execute to coordinate multiple agents [48]. For example, in a cooperative hunting task, co-agents (wolves) postulate a supraindividual "imagined we" with its own beliefs, desires, and intentions. Each wolf infers this agent's mental states by observing its own and others' actions, determines what the "imagined we" would do if rational, and performs its part of the collective plan. Results show that this agent-neutral "we" model improves coordination and produces more human-like cooperative behavior than individualistic strategies with shared rewards. A related notion of a "public belief", which depends on the actions taken by all the co-agents, has been adopted in a deep multi-agent reinforcement learning setting [55].

Various empirical studies already support the hypothesis that agent-neutral models may be necessary for achieving the coordination required for successful joint action, such as when cooking, playing, or moving house together [46,56,57]. A study of action preparation showed that individuals preparing for joint action benefitted from receiving advance information about the joint outcome to be produced (e.g., hand configuration formed by two individuals' hands), even if the individual contributions to the joint action (e.g., rotating one's hand to the left or right) remained unspecified [58]. In another study, the instruction to play a melody together rather than two melodies in parallel fundamentally changed how action-outcome relations were coded when individuals produced sounds [59]. Further studies support the proposal that action monitoring involves agent-neutral models. An EEG study of joint duet playing measured evoked response potentials (ERPs) in response to different types of errors inserted into the duet performance in real-time [60]. Individual errors that changed the nature of the joint outcome (the harmony between jointly produced tones) led to larger changes in error-related ERP components. A recent fMRI study identified a brain network that monitors errors during joint action [61]. Agent-neutral models may play a key role in learning [62] and imitating coordinated joint action, as when a novice couple dancing tango learns from an expert couple dancing tango [63,64]. Our previous research also provided the first indications that agent-neutral models may explain why successful joint outcomes induce a sense of vicarious agency in co-agents [33,65–67] and how a basic understanding of commitment emerges in joint action [13,68].

In sum, various computational modeling, theoretical, and empirical studies have recently addressed essential aspects of agent-neutral predictive models, suggesting they might be particularly effective in supporting joint

action. However, despite these advances, several challenges remain. The models discussed above incorporate essential aspects of agent-neutrality but are largely problem-specific. They do not yet capture realistic, multimodal joint actions involving both coordination and communication, and they have not been empirically tested. They should thus be seen as foundations for a novel approach to formalizing joint action rather than complete solutions. Moreover, key implications of agent-neutral predictive models still need to be understood and tested empirically. Under the assumption that agent-neutral predictive modeling follows the principles of predictive processing, successful joint action should efficiently minimize prediction errors about collective outcomes, and crucially, co-agents should represent and cooperate to minimize such prediction errors. Future studies should establish behavioral and neural correlates of agent-neutral predictive processing, for example, collective outcome predictions, during joint action, and test the hypothesis that co-agents represent and minimize them. A further hypothesis is that there might be a fundamental link between agent-neutral models and subjective experiences that accompany joint action, such as the sense of joint agency and the sense of commitment. A previous study assessed that the individual sense of agency is reduced during joint action, but can be restored at an implicit level when the self-other distinction and individual contributions are made salient [69]. Future studies could assess whether using agent-neutral models should lead co-agents to experience *joint agency* over action outcomes [10,70,71].

Second claim: Co-agents update their agent-neutral models interactively to better align them by shaping others' predictions through verbal and nonverbal communication.

Agent-neutral shared models must be regularly updated and aligned to adapt to the constantly changing dynamics of the world. We propose that co-agents use verbal and non-verbal communication to update and share agent-neutral models continuously. Communication dynamics create a continuous tension between *using* existing agent-neutral models versus *updating* them. When *using* existing agent-neutral models to control joint actions, co-agents should act and communicate in predictable and legible ways to not violate the agent-neutral model's predictions. However, when the co-agents' models are misaligned and when contextual conditions change, the models need to be *updated* to achieve a better alignment (Figure 2B). To achieve this "interactive alignment", co-agents need to surprise each other (i.e., take actions that contradict the predictions of other agents), where surprise serves as a method for exchanging new information and compelling to-agents to update and enrich their agent-neutral models, making them better able to minimize future surprises. For example, when driving in a convoy, one driver might take an unfamiliar route to avoid traffic (e.g., because she knows the neighborhood better). This decision could surprise the other drivers but also give them new information that will help update their shared area map. Interactive alignment also occurs during linguistic conversations, understood here as joint actions, on par with physical joint action. Speakers maintain a shared agent-neutral model – a shared contextual understanding of the situation – by relying on agreed-upon predictable utterances [72]. However, speakers often surprise listeners to enhance and update the shared understanding with new information that the listeners may not know. This new information leads the listener to update their internal contextual representation of the situation to better align with the speaker's model.

In keeping with our proposal, motor coordination and joint action studies show that pairs of people naturally use nonverbal, sensorimotor (or nonverbal) communication – and the strategy to momentarily surprise each other – to enhance their interaction [32]. For instance, when lifting or grasping objects together, co-agents adjust their movements to make their actions more predictable and their intentions more understandable to each other, especially when one agent has more information about the shared goal. Sensorimotor communication has been formalized in predictive processing as an effective mechanism to minimize others' uncertainty about the joint action and, more broadly, to align belief states, by temporarily surprising others [73]. Many of the mechanisms discussed above – agent-neutrality as a driving force for coordinating joint action and the importance of driving others' expectations and surprises to update the agent-neutral models – are also central in theoretical [14] and modeling accounts of natural language conversations [74]. Classical psycholinguistic theories focus on individual-based, modular processing of phonology, syntax, and semantics during verbal communication. In contrast, we view language as a joint action among two or more individuals that relies on agent-neutral, multi-subject alignment of meaning and goals. From this perspective, the basic unit for communication is two minds and brains rather than one: that of the speaker and that of the listener. Language is already a shared medium for conversation; each speaker enhances the conversation, improving the alignment with the other speaker by minimizing joint prediction errors. Remarkably, this simple process of reducing surprise while predicting the speaker's subsequent utterances has been used to effectively train large language models, such as GPT4 and Llama3. Recent papers using intracranial recordings of natural conversations showed that the human brain uses similar predictive mechanisms, like next-word prediction, also to process natural language in real-life conversations [75,76]. Subsequent research demonstrated that the embedding space acquired by large language models can be employed as a precise computational model of the

collective, context-heavy meaning space that connects the speaker's and listener's neural responses during free-flowing conversations [72].

In sum, various computational and empirical studies have assessed that both physical interactions and linguistic conversation, both conceived as joint actions, require a continuous effort of the co-agents to align and update agent-neutral models interactively, through nonverbal and verbal communication. Despite these advances, many facets of the interactive alignment of agent-neutral models remain to be investigated. Consider, for example, that during joint action and conversation, co-agents need to alternate between phases in which they rely upon a shared contextual representation of the situation (by acting in largely predictable ways) and when they update it through interactive alignment (by introducing surprising information). However, it is still unclear how people address the trade-off between achieving the joint action, which requires predictability, and aligning and updating agent-neutral models, which requires generating surprise. The theory of active inference formally accounts for a similar trade-off in individual cognition by positing that prediction error minimization during plan selection jointly optimizes a pragmatic imperative (exploiting known plans to achieve goals) and an epistemic imperative (exploring novel plans to reduce uncertainty about their outcomes) [38]. It is possible to speculate that the same formal framework can be extended to joint action, where the pragmatic imperative concerns achieving joint goals and the epistemic imperative involves transiently generating surprise to update and align agent-neutral models across individuals. In turn, the shift to interactive alignment might be driven by a perceived misalignment between the co-agents' agent-neutral models, which generates inconsistent prediction error minimization dynamics. However, these speculative hypotheses remain to be tested empirically.

Another open question is how co-agents negotiate conditions requiring updates to agent-neutral models. Our framework predicts that when expertise or knowledge is asymmetric, responsibility for alignment typically falls on the more knowledgeable agent, whose precise model generates larger prediction errors that readily signal misalignments. This agent is also better positioned to achieve realignment, consistent with findings that "leaders" use sensorimotor communication to support "followers." Less clear is how these principles apply without obvious asymmetries—for example, whether the co-agent detecting larger or earlier prediction errors assumes responsibility for alignment. In such cases, it also remains open how factors like power, hierarchy, and motivation shape the negotiation and updating of agent-neutral models. Finally, interactive alignment may also serve as a measure of commitment. Continuous efforts to align and update agent-neutral models can indicate agents' commitment to acting together, while the interactive nature of alignment suggests this commitment is mutual. Future studies could test whether agent-neutral models increase commitment to participate or remain engaged in joint action even without obligation [77,78], and whether this commitment depends on efforts to support alignment and on the degree of alignment achieved.

Third claim: Shared agent-neutral models support dynamical role allocation.

When working together toward a common goal, we often allocate specific roles to different individuals who then work on subgoals independently. For example, a team of three movers with the joint goal of moving a house can decide that two movers will load the truck while the third mover will pack the kitchen dishes. When cooking together, a couple might divide the work so that one person makes the salad while the other grills the fish. With repeated interactions, co-agents develop norms and conventions about role allocation that provide strong expectations about who does what [79,80]. At the same time, changing task demands require dynamic role allocation, as co-agents must adjust roles and subgoals. For instance, one agent can assist another in achieving their goal after completing their task.

Our framework proposes that dynamic role allocation and reallocation emerges naturally when conditions change, provided *agent-neutral* models are used to pursue the joint goal, since these models do not assume predefined roles. Using agent-neutral models allows co-agents to negotiate roles and collaborate on subgoals as the interaction evolves, guided by the principle of collective prediction error minimization: individuals can flexibly self-allocate and reallocate to roles where they can most effectively reduce collective prediction error (see Figure 2C). This process relies on continuous, reciprocal negotiation and exchange of informative signals, mirroring the interactive processing in our second claim. A drawback, however, is the cognitive and communicative cost, creating a trade-off between flexibility (keeping options open) and stability and predictability, which arise by forming conventions, which allow co-agents to reuse strategies to coordinate and predict, minimizing extra cognitive load. Active inference theory captures a similar trade-off in individual cognition when deciding whether to maintain or revise an existing predictive model of the situation. This trade-off is resolved by weighting the benefits of a novel model (accuracy) against the cost of changing one's mind

(complexity cost) [38,81]. Extending this idea, it is possible to speculate that maintaining or breaking a convention reflects an optimal trade-off between the benefits of addressing the situation more flexibly and the cognitive cost of changing collective expectations, a choice ultimately guided by the long-term minimization of prediction errors about collective outcomes. In turn, from an agent-neutral view, shifts toward, between, or away from roles and conventions should be guided by collective prediction error dynamics. However, these speculative hypotheses remain to be tested empirically.

There is currently a limited number of studies addressing multi-agent tasks requiring dynamic role allocation and the trade-offs implicit in them. One study showed that repeated interactions during joint planning tasks support the formation of stable strategies to decompose tasks into individual subgoals [82]. Interestingly, the same study shows that these strategies remain stable even when the conditions for their initial development change, suggesting that the advantage of sticking to existing social conventions [83] - and remaining predictable - could overcome the advantage of exploring novel and potentially more advantageous conventions. Furthermore, the study finds that in particular situations, such as when the task load is distributed unevenly, co-actors spontaneously allocate them to solve part of the co-actor's subgoals. Still, it does not systematically study these dynamical reallocations. Another study suggests that when allowed to select plans that split roles by precisely specifying individual subgoals versus specifying them more vaguely, thus leaving role allocation more open, participants preferred the flexibility of a vague plan when this was more advantageous according to a probabilistic model of joint reasoning [84]. The coordination and dynamic allocation of sub-tasks has received limited attention in computational modeling. One study examined role allocation in a multi-agent game where two or three co-agents collaborate to cook recipes quickly. Results show that by simultaneously inferring task responsibilities, small teams can rapidly align plans and coordinate work, avoiding interference [53]. The model also predicted human inferences about sub-task allocation. However, it does not allow co-agents to recognize completed tasks or flexibly reallocate to others. Moreover, the study does not test applicability to large-scale problems or investigate conditions and trade-offs in the emergence of social conventions during gameplay. Other research has successfully applied deep reinforcement learning to address large-scale, competitive team games that require co-agents to time their actions and distribute tasks and roles (e.g., attackers and defenders) [85]. However, these studies have not systematically analyzed the dynamics of role allocation or the specific features of the models that support it. The potential role of agent-neutral models in computational frameworks focusing on multi-agent setups with dynamic role allocation remains to be explored.

As only a few studies have explored our claim that shared agent-neutral models facilitate dynamic role allocation, many aspects of this claim still require investigation. A prediction stemming from our proposal is that co-agents (having the appropriate skills) will flexibly allocate themselves to any roles and subgoals a joint action offers; and this dynamic allocation and re-allocation of roles should depend on collective prediction error dynamics. Another prediction stemming from our proposal is that agent-neutral models should enable co-agents to focus on the collective goal and on role-specific subgoals simultaneously. If this is the case, co-agents should simultaneously monitor and reduce prediction errors related to multiple goals and allocate more attention or precision to specific goals based on the current task demands. This prediction could be addressed by developing novel scenarios and techniques to dissect and monitor simultaneously different types of prediction errors during joint action. Finally, another prediction from our proposal is that when co-agents prioritize group goals over individual goals, they should be more committed to working together. Likewise, co-agents should accept an unbalanced (perhaps even unfair) distribution of effort between individuals if the collective goal is prioritized. Future studies that systematically vary the costs and benefits of individual and collective goals might help test these predictions.

Discussion

In this paper, we advanced a novel integrated predictive processing framework for joint action and communication based upon agent-neutral predictive models as a novel basis for understanding the individual basis of human sociality. The framework builds on principles of predictive processing that have led to fundamental advances in understanding the dynamics of individual perception and action and the nature of language processing [36–38]. We have highlighted three key claims from our new agent-neutral perspective, discussed the empirical evidence supporting them, and suggested possible future experiments to test these claims.. Instead of concentrating on predicting and controlling the outcomes of individual actions—which remains the prevailing perspective in individual motor control and social cognition—we advocate for predicting and controlling joint outcomes. This shift in focus explains how individuals can learn to coordinate with others to achieve joint outcomes effectively and why there is little cognitive effort involved in conversing

and aligning our thoughts with others during everyday conversations. Our emphasis on agent-neutral models also provides a new perspective for understanding why and when individuals exchange new, surprising information to align their mental states during joint action and communication. In particular, agent-neutral models enable individuals to identify information of added value (surprising) for others, enhance the alignment of goals, and propagate beliefs across agents. Agent-neutral models create the scaffold for negotiating and aligning relevant goals and beliefs during natural language conversations, to allocate roles during joint actions with multiple subgoals and to develop conventions. These examples emphasize that our proposal could imply a radical shift in how social cognition, cooperation, and communication are studied. However, while existing evidence already provides some support for our claims, many novel predictions stemming from them remain to be investigated.

In the remainder of the paper, we address several open questions raised by the framework: what is shared in agent-neutral representations; whether agent-neutral and agent-specific models can coexist; whether linguistic communication should be seen merely as a mechanism supporting joint action or as a form of joint action itself; how does our framework relate to previous accounts of social cognition; and what its potential impact may be across domains.

What is shared in agent-neutral representations?

We have emphasized the importance of shared agent-neutral representations. Now that we have outlined our three main claims, it is useful to revisit what we mean by "sharing." During social interactions, models often align in various ways. For example, in a collaborative task like building a tower (Figure 2), co-agents typically agree on the available pieces and which to use, based on a common understanding of the goal. Predictable actions, such as adding a red brick to a red tower, arise from these shared agent-neutral models. The same applies to language, which adds context that allows anticipation of phrases from previous conversations and the situation. These examples show that much is already shared between co-actors even without explicit coordination. In cooperative settings, shared agent-neutral models generate coherent action trajectories and predictions for all participants. Similarly, in competitive environments, such as soccer, players exploit shared expectations to outmaneuver opponents. Agent-neutral models are however not always fully shared; co-actors may have differing representations of joint goals or tasks. Successful collaboration often relies on achieving shared understanding, but it doesn't necessarily require maintaining separate models for "my" versus "my partner's" agent-neutral model. Instead, prediction errors from one's own model can identify and correct misalignments, see [44,45] for examples.

Can agent-neutral and agent-specific models co-exist?

The present framework does not preclude more complex strategies for monitoring and achieving shared understanding. Such strategies may involve separate representations for one's own and others' beliefs, typically linked to "mind-reading." The central claim, however, is that these mechanisms are not universally necessary and may play a less prominent role than often assumed; their involvement should be demonstrated empirically, not presupposed. Likewise, while this framework emphasizes agent-neutral models in joint action, it does not exclude agent-specific ones. Individuals may rely on agent-specific representations in parallel with, or sometimes in place of, agent-neutral ones, depending on context. Theoretical and computational models illustrate how both types can coexist hierarchically—for example, in joint actions requiring subtask allocation, agent-specific subgoals may be integrated within an overarching agent-neutral model [45,46,86]. Importantly, agent-specific plans can also be derived from agent-neutral models, as in the computational models reviewed, where individual plans are executed but grounded in agent-neutral structures (e.g., an "imagined we"). Thus, agent-specific models can guide individual actions in ways sensitive to collective outcomes. What remains unclear is under which conditions, if any, agent-neutral models are precluded.

It might be argued that agent-specific models are the default when agent-neutral models are not yet available. Indeed, like any predictive model, agent-neutral ones require experience to develop. Yet there is little reason to view their acquisition as secondary; they may reflect a natural and pervasive mode of social learning throughout development. Specialized agent-neutral models for activities like team sports or musical ensembles may require extensive training, but more general-purpose models for everyday interactions—conversation or simple cooperation—are likely acquired early and generalized flexibly. A key advantage is that their use does not presuppose familiarity with a specific partner, though such experience can fine-tune them. This aligns with proposals of an "interaction engine" scaffolding social coordination [3]. From this perspective, adaptation of agent-neutral models operates across two timescales. At a fast timescale, interactive alignment and sharing correspond to updating posterior beliefs in a Bayesian framework. At slower timescales of learning and

development, agent-neutral models improve through accumulated experience, corresponding to parameter updating. The latter is especially relevant in novel joint action scenarios, such as when co-agents are paired for the first time with differing skills (e.g., expert vs. novice).

Linguistic communication: a mechanism supporting joint action or a joint action in itself?

Another recurrent theme of our framework is the role of agent-neutral models in both physical action and linguistic communication. Notably, communication in joint action can be conceptualized in two ways. One view treats language as a tool for aligning co-agents' models of physical action, thereby improving coordination. The other, which we adopt, considers linguistic communication as a joint action in its own right. From this perspective, communication involves agent-neutral models and joint goals distinct from those driving physical action. These goals are often epistemic—achieving shared understanding, acquiring information, and co-constructing narratives of reality—but may also be prosocial, such as seeking affiliation. This contrasts with the pragmatic goals emphasized in physical joint action, like lifting an object or building a structure. Although epistemic goals are sometimes regarded as secondary, they are powerful drivers of human behavior and perhaps of other social species. The pursuit of knowledge, shared narratives, and cultural construction reflects coordinated efforts often mediated by language. These efforts operate not only at the short timescales of conversations and moment-to-moment coordination, but also at the longer timescales of social and cultural evolution.

How does our novel framework relate to previous accounts of social cognition?

From a theoretical perspective, our proposal aligns well with recent efforts to shift social cognition from a first-person to a second-person perspective, emphasizing the crucial role of interactive dynamics within it [87,88]. Previous theoretical proposals have emphasized related concepts that align closely with our framework, including standard ground formation [7], shared representations [50], "we representations" [87], and the influence of joint goals in guiding predictive processes during motor interactions [19,89]. While our framework is conceptually related to these accounts, it is grounded in predictive processing and introduces novel elements, most notably, the central role of prediction error mechanisms in monitoring the combined outcomes of co-actors. The hierarchical predictive architecture for joint action proposed by [86] also bears similarities to our approach, particularly in that its highest level can be construed as agent-neutral. This feature is shared with some of the computational models discussed earlier. However, unlike our first claim, their model posits that agent-neutral processing is restricted to the top level, with lower levels generating separate predictions for self and other. This reliance on agent-specific processing at lower levels may limit the scalability of this proposal (as well as of related ones that emphasize agent-specific models) in contexts involving larger groups or complex coordination. This problem is also present in other theoretical and computational accounts that assume multiple agent-specific models, although it can be partly mitigated by considering a single agent-specific model for the self and a collective model for the (average) behavior of any number of co-agents [54]. Moreover, while our second claim emphasizes the role of predictive mechanisms, based upon strategic surprise generation and minimization, not only in action coordination but also in the interactive construction and updating of agent-neutral models, these interactive and epistemic dimensions of predictive processing are not fully discussed in the previous theoretical proposals considered here.

The predictive processing framework proposed in this paper does not contradict but instead aims to both synthesize and extend these existing accounts proposed in cognitive science and neuroscience, along several dimensions. It contextualizes previous theories emphasizing shared intentionality, joint payoffs, joint experiences, or joint task representations, within a broader framework focused on the predictive brain. Besides the importance of integration, this move has the advantage of linking more tightly theories of joint action and communication with principles of brain processing. Another innovative aspect of our proposal is the use of the same computational principles to address joint action coordination and interactive alignment during conversations in natural language. This has great potential to lead to the closer integration of disparate research findings. This will include new answers to how communication supports collaboration and how the need for communication arises in joint action. Linking subjective experience to computational principles can significantly enhance our understanding of what makes people feel they are gaining or losing control during joint action. This could be highly relevant in formalizing joint action dynamics more precisely and resolving urgent questions concerning the interaction between humans and artificial agents. Some preliminary steps in this direction have been made with the aid of computational models that incorporate aspects of predictive processing and agent-neutrality. Still, much remains to be done to make these models or their extension better able to deal with the complexities of joint action in realistic contexts.

Furthermore, our framework does not focus on a single type of joint action but aims to capture its full complexity, from short-term, small-team interactions to large-scale, long-lasting ones, encompassing symmetric and asymmetric roles and knowledge, physical and conversational interactions, and pragmatic and epistemic goals. Addressing this variety requires mechanisms that ensure scalability, flexibility, and a balance between dynamicity and persistence. Our three claims illustrate how interactive predictive processing provides these benefits. The first claim emphasizes agent-neutral over agent-specific models and minimization of prediction errors about collective outcomes, allowing efficient scaling to multi-agent interactions without intractable computational costs. The second claim highlights flexibility: constructing and aligning agent-neutral models interactively allows the same principles to model symmetric and asymmetric joint actions, including leader–follower dynamics or expert–novice interactions. For example, temporarily surprising others to shape predictions can apply to both physical actions and conversations. The third claim addresses trade-offs between dynamicity (e.g., dynamic role allocation) and persistence (e.g., emergence of conventions). Agent-neutral models support flexible role assignment but require solving role allocation during interaction, which may incur cognitive or communicative costs. Once stable solutions emerge, they can be reused, stabilizing conventions over time. Together, these predictive processing mechanisms explain why joint action can be scalable, flexible, and balanced between dynamicity and persistence.

What is the potential impact of our novel framework across domains?

Given these premises, our new framework has enormous potential to impact our understanding of social interaction as it unfolds in the real world and help us understand (for example) the collaborative dynamics of cooks in a restaurant, factory workers, or medical teams performing surgery. It can be applied to a broad range of interactive settings, including ensemble music and dance [90], education [91,92], psychotherapy [93], team sports, and multi-agent AI settings, in which co-agents take complementary roles and learn from each other [94,95]. In educational contexts, having an agent-neutral computational model for everyday conversations may lay the foundation for translational work in improving teacher-student communication. In health-related contexts, our agent-neutral model offers significant potential for advancing our understanding of typical and atypical patterns of social communication and language comprehension and production, as seen in conditions such as autism spectrum disorder (ASD) or social (pragmatic) communication disorder. Furthermore, by studying the structure and flow of joint action during everyday conversations, researchers can gain insights into language production processes and the organization of thoughts into coherent speech. This knowledge can be applied to design therapeutic approaches and technologies that support individuals with speech production challenges, such as those experiencing aphasia, dysarthria, or stuttering.

From a more technological perspective, understanding how agent-neutral predictive models support human-human joint action might help solve practical problems that arise in interactions between humans and artificial systems, hence paving the way for novel technologies to advance human-robot interaction. A recent trend in robot learning applies the foundation model paradigm, successful in domains like large language models. Here, robot foundation models are trained on large, curated datasets, often minimizing autonomous exploration and social interaction, especially early in training. While this addresses the high cost and complexity of real-world interaction, it contrasts with biological development, where interaction and exploration are essential. This approach has advanced robotics, but it remains unclear whether it can produce robots capable of human-like interaction. If agent-neutral models are foundational to human joint action, current methods may need complementary training objectives to foster their emergence. This could be achieved through embodied interaction and shared experience, including collaboration with robots and humans. In other words, learning to interact may better support agent-neutral representations than relying solely on static datasets. It also remains to be determined whether robotic agent-neutral models can be acquired autonomously via joint reward maximization [85], or whether additional inductive biases are needed for robots to construct internal models of their own and others' "minds" as cooperative agents [96].

Conclusions

Summing up, we advance a new perspective for understanding the individual basis of human sociality, grounded in an interactive view of how predictive processing operates. Previous research has already shown that the body movements and brain dynamics of people engaged in joint actions tend to synchronize, align, and coordinate in various ways. Here, we propose that this is an index of the fundamentally interactive nature of our predictive processing mechanisms. When engaged in joint action, we adopt agent-neutral models that allow us to predict collective outcomes (i.e., the combined results of joint actions, regardless of who carries them out), calculate prediction errors about these collective outcomes, and engage in sensorimotor

communication to better align and share predictive models across individuals. This interactive form of predictive processing—or interactive inference—implies that the intentionality and agency of action must be attributed to the dyad or group of individuals. Over the long term, engaging in interactive inference may have supported the development of human culture, prosocial behavior, and emotions, and the construction of shared world models—that is, a largely shared understanding of the physical and social reality we inhabit, extending beyond what we could perceive or achieve as isolated individuals. We infer together the worlds we live in and the courses of action that change them to realize our goals.

Acknowledgements

This research received funding from the European Research Council under the Grant Agreement No. 820213 (ThinkAhead) to GP, the Italian National Recovery and Resilience Plan (NRRP), M4C2, funded by the European Union – NextGenerationEU (Project IR0000011, CUP B51E22000150006, "EBRAINS-Italy"; Project PE0000013, "FAIR"; Project PE0000006, "MNESYS") to GP, and the Ministry of University and Research, PRIN PNRR P20224FESY and PRIN 20229Z7M8N to GP. EP was supported by the French National Research Agency (ANR-10-IDEX-0001-02 PSL* and ANR-17-EURE-0017 FrontCog). UH was supported by NIH/NIDCD (1R01DC022534). We used a generative AI model to correct typographical errors and edit language for clarity and as an aid to sketch Figure 1.

References

1. Brownell CA. 2011 Early developments in joint action. *Review of philosophy and psychology* **2**, 193–211.
2. Henrich J. 2017 *The Secret of Our Success. How our collective intelligence has helped us to evolve and prosper*. Princeton University Press. See https://press.princeton.edu/books/paperback/9780691178431/the-secret-of-our-success.
3. Levinson SC. 2006 On the human 'interaction engine'. In *Roots of human sociality: Culture, cognition and interaction* (eds NJ Enfield, SC Levinson), pp. 39–69. Oxford: Berg.
4. Tomasello M, Carpenter M, Call J, Behne T, Moll H. 2005 Understanding and sharing intentions: the origins of cultural cognition. *Behav Brain Sci* **28**, 675–91; discussion 691-735. (doi:10.1017/S0140525X05000129)
5. Tuomela R. 2007 *The Philosophy of Sociality: The Shared Point of View*. Oxford University Press.
6. Gilbert M. 2013 *Joint Commitment: How We Make the Social World*. Oxford University Press. (doi:10.1093/acprof:oso/9780199970148.001.0001)
7. Clark HH. 1996 *Using language*. Cambridge university press.
8. Sperber D. 1996 Explaining Culture: a naturalistic approach.
9. Tomasello M. 2009 *Why we cooperate*. MIT press.
10. Loehr JD. 2022 The sense of agency in joint action: An integrative review. *Psychonomic Bulletin & Review* **29**, 1089–1117.
11. Pacherie E. 2014 How does it feel to act together? *Phenomenology and the cognitive sciences* **13**, 25–46.
12. Michael J, Pacherie E. 2015 On commitments and other uncertainty reduction tools in joint action. *Journal of Social Ontology* **1**, 89–120.
13. Michael J, Sebanz N. 2016 The sense of commitment: A minimal approach. *Frontiers in psychology* **6**, 162497.
14. Garrod S, Pickering MJ. 2009 Joint action, interactive alignment, and dialog. *Topics in Cognitive Science* **1**, 292–304.
15. Hasson U, Ghazanfar AA, Galantucci B, Garrod S, Keysers C. 2012 Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends in cognitive sciences* **16**, 114–121.
16. Steels L. 2003 Evolving grounded communication for robots. *Trends in cognitive sciences* **7**, 308–312.
17. Zada Z *et al.* 2023 A shared linguistic space for transmitting our thoughts from brain to brain in natural conversations. *bioRxiv*
18. Bratman ME. 2013 *Shared agency: A planning theory of acting together*. Oxford University Press.
19. Butterfill SA, Sinigaglia C. 2023 Towards a Mechanistically Neutral Account of Acting Jointly: The Notion of a Collective Goal. *Mind* **132**, 1–29. (doi:10.1093/mind/fzab096)
20. Gilbert M. 1992 *On social facts*. Princeton University Press.
21. Wilson D, Sperber D. 2006 Relevance theory. *The handbook of pragmatics* , 606–632.
22. Bacharach M. 2006 *Beyond individual choice*. Princeton, NJ: Princeton Univ. Press.

23. Sugden R. 2003 The logic of team reasoning. *Philosophical explorations* **6**, 165–181.
24. Hasson U, Frith CD. 2016 Mirroring and beyond: coupled dynamics as a generalized framework for modelling social interactions. *Phil. Trans. R. Soc. B* **371**, 20150366. (doi:10.1098/rstb.2015.0366)
25. Marsh KL, Richardson MJ, Schmidt RC. 2009 Social connection through joint action and interpersonal coordination. *Topics in cognitive science* **1**, 320–339.
26. Bekkering H, De Bruijn ER, Cuijpers RH, Newman-Norlund R, Van Schie HT, Meulenbroek R. 2009 Joint action: Neurocognitive mechanisms supporting human interaction. *Topics in Cognitive Science* **1**, 340–352.
27. Gallese V, Keysers C, Rizzolatti G. 2004 A unifying view of the basis of social cognition. *Trends Cogn Sci* **8**, 396–403. (doi:10.1016/j.tics.2004.07.002)
28. Pickering MJ, Garrod S. 2013 An integrated theory of language production and comprehension. *Behavioral and brain sciences* **36**, 329–347.
29. Misyak JB, Chater N. 2014 Virtual bargaining: a theory of social decision-making. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, 20130487.
30. Wolpert DM, Doya K, Kawato M. 2003 A unifying computational framework for motor control and social interaction. *Philos Trans R Soc Lond B Biol Sci* **358**, 593–602. (doi:10.1098/rstb.2002.1238)
31. Sebanz N, Knoblich G. 2009 Prediction in joint action: What, when, and where. *Topics in cognitive science* **1**, 353–367.
32. Pezzulo G, Donnarumma F, Dindo H, D'Ausilio A, Konvalinka I, Castelfranchi C. 2019 The body talks: Sensorimotor communication and its brain and kinematic signatures. *Physics of Life Reviews* **28**, 1–21. (doi:10.1016/j.plrev.2018.06.014)
33. Dewey JA, Pacherie E, Knoblich G. 2014 The phenomenology of controlling a moving object with another person. *Cognition* **132**, 383–397.
34. Dewey JA, Knoblich G. 2014 Do implicit and explicit measures of the sense of agency measure the same thing? *PloS one* **9**, e110118.
35. Obhi SS, Hall P. 2011 Sense of agency and intentional binding in joint action. *Experimental brain research* **211**, 655–662.
36. Clark A. 2015 *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press, Incorporated.
37. Hohwy J. 2013 *The predictive mind*. Oxford University Press.
38. Parr T, Pezzulo G, Friston KJ. 2022 *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. Cambridge, MA, USA: MIT Press.
39. Baker CL, Saxe R, Tenenbaum JB. 2009 Action understanding as inverse planning. *Cognition* **113**, 329–349. (doi:10.1016/j.cognition.2009.07.005)
40. Demiris Y, Khadhouri B. 2005 Hierarchical Attentive Multiple Models for Execution and Recognition (HAMMER). *Robotics and Autonomous Systems Journal* **54**, 361–369.
41. Dindo H, Zambuto D, Pezzulo G. 2011 Motor simulation via coupled internal models using sequential Monte Carlo. In *Proceedings of IJCAI 2011*, pp. 2113–2119.
42. Ramirez M, Geffner H. 2010 Probabilistic Plan Recognition using off-the-shelf Classical Planners. In *Proc. AAAI-10*, Atlanta, USA.
43. Devaine M, Hollard G, Daunizeau J. 2014 The social Bayesian brain: does mentalizing make a difference when we learn? *PLoS computational biology* **10**, e1003992.
44. Friston K, Frith C. 2015 A Duet for one. *Consciousness and cognition* **36**, 390–405.
45. Maisto D, Donnarumma F, Pezzulo G. 2023 Interactive Inference: A Multi-Agent Model of Cooperative Joint Actions. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* , 1–12. (doi:10.1109/TSMC.2023.3312585)
46. Masumoto J, Inui N. 2015 Motor control hierarchy in joint action that involves bimanual force production. *J Neurophysiol* **113**, 3736–3743. (doi:10.1152/jn.00313.2015)
47. McClelland K. 2004 The collective control of perceptions: constructing order from conflict. *International Journal of Human-Computer Studies* **60**, 65–99. (doi:10.1016/j.ijhcs.2003.08.003)
48. Tang N, Gong S, Zhao M, Gu C, Zhou J, Shen M, Gao T. 2022 Exploring an imagined "we" in human collective hunting: Joint commitment within shared intentionality. In *Proceedings of the annual meeting of the cognitive science society*,
49. Sebanz N, Bekkering H, Knoblich G. 2006 Joint action: bodies and minds moving together. *Trends in cognitive sciences* **10**, 70–76.
50. Sebanz N, Knoblich G. 2021 Progress in joint-action research. *Current Directions in Psychological Science* **30**, 138–143.
51. Pezzulo G, Iodice P, Ferraina S, Kessler K. 2013 Shared action spaces: a basis function framework for

social re-calibration of sensorimotor representations supporting joint action. *Frontiers in human neuroscience* **7**, 800.
52. Shum M, Kleiman-Weiner M, Littman ML, Tenenbaum JB. 2019 Theory of minds: understanding behavior in groups through inverse planning. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence*, pp. 6163–6370. Honolulu, Hawaii, USA: AAAI Press. (doi:10.1609/aaai.v33i01.33016163)
53. Wu SA, Wang RE, Evans JA, Tenenbaum JB, Parkes DC, Kleiman-Weiner M. 2021 Too Many Cooks: Bayesian Inference for Coordinating Multi-Agent Collaboration. *Topics in Cognitive Science* **13**, 414–432. (doi:10.1111/tops.12525)
54. Khalvati K, Park SA, Mirbagheri S, Philippe R, Sestito M, Dreher J-C, Rao RPN. 2019 Modeling other minds: Bayesian inference explains human choices in group decision-making. *Science Advances* **5**, eaax8783. (doi:10.1126/sciadv.aax8783)
55. Foerster J, Song F, Hughes E, Burch N, Dunning I, Whiteson S, Botvinick M, Bowling M. 2019 Bayesian action decoder for deep multi-agent reinforcement learning. In *International Conference on Machine Learning*, pp. 1942–1951. PMLR.
56. Noy L, Dekel E, Alon U. 2011 The mirror game as a paradigm for studying the dynamics of two people improvising motion together. *Proceedings of the National Academy of Sciences* **108**, 20947–20952.
57. Vesper C, Butterfill S, Knoblich G, Sebanz N. 2010 A minimal architecture for joint action. *Neural Networks* **23**, 998–1003.
58. Kourtis D, Woźniak M, Sebanz N, Knoblich G. 2019 Evidence for we-representations during joint action planning. *Neuropsychologia* **131**, 73–83.
59. Sacheli LM, Arcangeli E, Paulesu E. 2018 Evidence for a dyadic motor plan in joint action. *Scientific reports* **8**, 5027.
60. Loehr JD, Kourtis D, Vesper C, Sebanz N, Knoblich G. 2013 Monitoring individual and joint action outcomes in duet music performance. *Journal of cognitive neuroscience* **25**, 1049–1061.
61. Sacheli LM, Musco MA, Zazzera E, Banfi G, Paulesu E. 2022 How shared goals shape action monitoring. *Cerebral Cortex* **32**, 4934–4951.
62. Marschner M, Dignath D, Knoblich G. 2024 Me or we? Action-outcome learning in synchronous joint action. *Cognition* **247**, 105785.
63. Ramenzoni VC, Sebanz N, Knoblich G. 2015 Synchronous imitation of continuous action sequences: The role of spatial and topological mapping. *Journal of Experimental Psychology: Human Perception and Performance* **41**, 1209.
64. Tsai JC-C, Sebanz N, Knoblich G. 2011 The GROOP effect: Groups mimic group actions. *Cognition* **118**, 135–140.
65. Bars SL, Bourgeois-Gironde S, Wyart V, Sari I, Pacherie E, Chambon V. 2022 Motor Coordination and Strategic Cooperation in Joint Action. *Psychol Sci* **33**, 736–751. (doi:10.1177/09567976211053275)
66. Le Bars S, Devaux A, Nevidal T, Chambon V, Pacherie E. 2020 Agents' pivotality and reward fairness modulate sense of agency in cooperative joint action. *Cognition* **195**, 104117. (doi:10.1016/j.cognition.2019.104117)
67. van der Wel RP, Sebanz N, Knoblich G. 2012 The sense of agency during skill learning in individuals and dyads. *Consciousness and cognition* **21**, 1267–1279.
68. Michael J, Sebanz N, Knoblich G. 2016 Observing joint action: Coordination creates commitment. *Cognition* **157**, 106–113.
69. Zapparoli L, Mariano M, Sacheli LM, Berni T, Negrone C, Toneatto C, Paulesu E. 2024 Self-other distinction modulates the sense of self-agency during joint actions. *Sci Rep* **14**, 30055. (doi:10.1038/s41598-024-80880-7)
70. Pacherie E. 2012 The phenomenology of joint action: Self-agency versus joint agency.
71. Zapparoli L, Paulesu E, Mariano M, Ravani A, Sacheli LM. 2022 The sense of agency in joint actions: A theory-driven meta-analysis. *Cortex* **148**, 99–120.
72. Zada Z *et al.* 2024 A shared model-based linguistic space for transmitting our thoughts from brain to brain in natural conversations. *Neuron*
73. Pezzulo G, Donnarumma F, Dindo H. 2013 Human Sensorimotor Communication: A Theory of Signaling in Online Social Interactions. *PLoS ONE* **8**, e79876. (doi:10.1371/journal.pone.0079876)
74. Ying L, Jha K, Aarya S, Tenenbaum JB, Torralba A, Shu T. 2024 GOMA: Proactive Embodied Cooperative Communication via Goal-Oriented Mental Alignment. *arXiv preprint arXiv:2403.11075*
75. Goldstein A *et al.* 2022 Shared computational principles for language processing in humans and deep language models. *Nature neuroscience* **25**, 369–380.

76. Goldstein A *et al.* 2023 Deep speech-to-text models capture the neural basis of spontaneous speech in everyday conversations. *bioRxiv* , 2023–06.
77. Fernández-Castro V, Pacherie E. 2023 Commitments and the sense of joint agency. *Mind & Language* **38**, 889–906. (doi:10.1111/mila.12433)
78. Michael J. 2022 *The philosophy and psychology of commitment*. Taylor & Francis.
79. Bicchieri C. 2005 *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.
80. Conte R, Castelfranchi C. 1995 *Cognitive and Social Action*. London, UK: University College London.
81. Zenon A, Solopchuk O, Pezzulo G. 2017 An information-theoretic perspective on the costs of cognition. *bioRxiv* , 208280. (doi:10.1101/208280)
82. Gordon J, Knoblich G, Pezzulo G. 2023 Strategic task decomposition in joint action. *Cognitive Science* **47**, e13316.
83. Theriault JE, Young L, Barrett LF. 2021 The sense of should: A biologically-based framework for modeling social pressure. *Physics of Life Reviews* **36**, 100–136.
84. Yu D, Thompson B. 2025 Adaptive use of vagueness to coordinate joint action. *PsyArXiv Preprints*
85. Jaderberg M *et al.* 2019 Human-level performance in 3D multiplayer games with population-based reinforcement learning. *Science* **364**, 859–865. (doi:10.1126/science.aau6249)
86. Pesquita A, Whitwell RL, Enns JT. 2018 Predictive joint-action model: A hierarchical predictive approach to human cooperation. *Psychon Bull Rev* **25**, 1751–1769. (doi:10.3758/s13423-017-1393-6)
87. Gallotti M, Frith CD. 2013 Social cognition in the we-mode. *Trends Cogn. Sci. (Regul. Ed.)* **17**, 160–165. (doi:10.1016/j.tics.2013.02.002)
88. Schilbach L, Timmermans B, Reddy V, Costall A, Bente G, Schlicht T, Vogeley K. 2013 Toward a second-person neuroscience. *Behavioral and brain sciences* **36**, 393–414.
89. Butterfill S. 2012 Joint Action and Development. *The Philosophical Quarterly* **62**, 23–47. (doi:10.1111/j.1467-9213.2011.00005.x)
90. Abalde SF, Rigby A, Keller PE, Novembre G. 2024 A framework for joint music making: behavioral findings, neural processes, and computational models. *Neuroscience & Biobehavioral Reviews* , 105816.
91. Meshulam M, Hasenfratz L, Hillman H, Liu Y-F, Nguyen M, Norman KA, Hasson U. 2021 Neural alignment predicts learning outcomes in students taking an introduction to computer science course. *Nat Commun* **12**, 1922. (doi:10.1038/s41467-021-22202-3)
92. Nguyen M, Chang A, Micciche E, Meshulam M, Nastase SA, Hasson U. 2022 Teacher–student neural coupling during teaching and learning. *Social Cognitive and Affective Neuroscience* **17**, 367–376. (doi:10.1093/scan/nsab103)
93. Zilcha-Mano S. 2024 How getting in sync is curative: Insights gained from research in psychotherapy. *Psychological Review*
94. Gordon J, Maselli A, Lancia GL, Thiery T, Cisek P, Pezzulo G. 2021 The road towards understanding embodied decisions. *Neuroscience & Biobehavioral Reviews* (doi:10.1016/j.neubiorev.2021.09.034)
95. Maselli A, Gordon J, Eluchans M, Lancia GL, Thiery T, Moretti R, Cisek P, Pezzulo G. 2023 Beyond simple laboratory studies: Developing sophisticated models to study rich behavior. *Phys Life Rev* **46**, 220–244. (doi:10.1016/j.plrev.2023.07.006)
96. Pezzulo G, Parr T, Friston KJ. 2025 Shared worlds, shared minds: Strategies to develop physically and socially embedded AI. *EMBO reports* , 1–6. (doi:10.1038/s44319-025-00549-8)