Interactive Inference: A Multi-Agent Model of Cooperative Joint Actions

Domenico Maisto, Francesco Donnarumma, and Giovanni Pezzulo

Abstract—**We advance a novel computational model of multi-agent, cooperative joint actions that is grounded in the cognitive framework of active inference. The model assumes that to solve a joint task, such as pressing together a red or blue button, two (or more) agents engage in a process of interactive inference. Each agent maintains probabilistic beliefs about the joint goal (e.g., Should we press the red or blue button?) and updates them by observing the other agent's movements, while in turn selecting movements that make his own intentions legible and easy to infer by the other agent (i.e., sensorimotor communication). Over time, the interactive inference aligns both the beliefs and the behavioral strategies of the agents, hence ensuring the success of the joint action. We exemplify the functioning of the model in two simulations. The first simulation illustrates a "leaderless" joint action. It shows that when two agents lack a strong preference about their joint task goal, they jointly infer it by observing each other's movements. In turn, this helps the interactive alignment of their beliefs and behavioral strategies. The second simulation illustrates a "leader–follower" joint action. It shows that when one agent ("leader") knows the true joint goal, it uses sensorimotor communication to help the other agent ("follower") infer it, even if doing this requires selecting a more costly individual plan. These simulations illustrate that interactive inference supports successful multi-agent joint actions and reproduces key cognitive and behavioral dynamics of "leaderless" and "leader–follower" joint actions observed in human–human experiments. In sum, interactive inference provides a cognitively inspired, formal framework to realize cooperative joint actions and consensus in multi-agent systems.**

Index Terms—**Active inference, consensus, joint action, multi-agent systems (MASs), sensorimotor communication, shared knowledge, social interaction.**

Manuscript received 25 May 2023; accepted 30 August 2023. Date of publication 20 October 2023; date of current version 17 January 2024. The work of Francesco Donnarumma was supported in part by MUR–PRIN2020 under Grant 2020529PCP. The work of Giovanni Pezzulo was supported in part by the EU Horizon 2020 Framework Programme for Research and Innovation through Human Brain Project SGA3 under the Specific Grant Agreement 945539 and through the TAILOR under Grant 952215; in part by the European Research Council through ThinkAhead under Grant 820213; and in part by PNRR MUR under Grant PE0000013-FAIR and Grant IR00011-EBRAINS-Italy. This article was recommended by Associate Editor F. Chiclana. *(Corresponding author: Giovanni Pezzulo.)*

The authors are with the Institute of Cognitive Sciences and Technologies, National Research Council, 00185 Rome, Italy (e-mail: domenico.maisto@cnr.it; francesco.donnarumma@istc.cnr.it; giovanni.pezzulo@istc.cnr.it).

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/TSMC.2023.3312585.

Digital Object Identifier 10.1109/TSMC.2023.3312585

I. Introduction

A CENTRAL challenge of multi-agent systems (MASs) is coordinating the actions of multiple autonomous agents in time and space, to accomplish cooperative tasks and achieve joint goals [1], [2]. Developing successful MASs requires addressing controllability challenges [3], [4] and dealing with synchronization control [5] formation control [6], task allocation [7], and consensus formation [8], [9], [10].

Research in cognitive science may provide guiding principles to address the above challenges, by identifying the cognitive strategies that groups of individuals use to successfully interact with each other and to make collective decisions [11], [12], [13], [14]. An extensive body of research studied how two or more people coordinate their actions in time and space during cooperative (human–human) joint actions, such as when performing team sports, dancing or lifting something together [15], [16]. These studies have shown that successful joint actions engage various cognitive mechanisms, whose level of sophistication plausibly depends on task complexity. The simplest forms of coordination and imitation in pairs or groups of individuals, such as the joint execution of rhythmic patterns, might not require sophisticated cognitive processing, but could use simple mechanisms of behavioral synchronization—perhaps based on coupled dynamical systems, analogous to the synchronization of coupled pendulums [17]. However, more sophisticated types of joint actions go beyond the mere alignment of behavior. For example, some joint actions require making decisions together, e.g., the decision about where to place a table that we are lifting together. These sophisticated forms of joint actions and joint decisions might benefit from cognitive mechanisms for mutual prediction, mental state inference, sensorimotor communication and shared task representations [16], [18], [19]. The cognitive mechanisms supporting joint action have been probed by numerous experiments [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], sometimes with the aid of conceptual [31], computational [32], [33], [34], [35], [36], [37], [38], [39], [40], and robotic [41], [42], [43], [44] models. Despite this progress, there is a paucity of models that implement advanced cognitive abilities, such as the inference of others' plans and the alignment of task knowledge across group members. Furthermore, we still lack a formal theory that explains the cognitive mechanisms of joint actions from first principles; for example, from the perspective of a generic inference or optimization scheme.

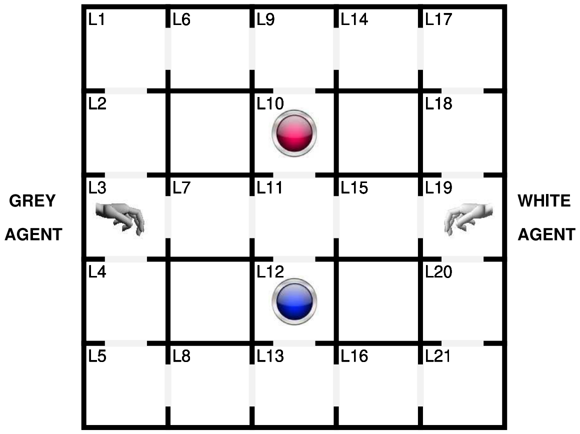

Fig. 1. Schematic illustration of the "joint maze" task. The two (gray and white) agents are represented as two hands. Their initial positions are L3 (gray) and L19 (white). Their possible goal locations are in blue (L12) and red (L10). The agents can navigate in the maze, by following the open paths, but cannot go through walls).

We advance an innovative framework for cooperative joint action and consensus in MASs, inspired by the cognitive framework of active inference. Active inference is a normative theory that describes the brain as a prediction machine, which learns an internal (generative) model of the statistical regularities of the environment—including the statistics of social interactions—and uses it to generate predictions that guide perceptual processing and action planning [45]. Here, we use the predictive and inferential mechanisms of active inference to implement sophisticated forms of joint action in dyads of interacting agents. The model presented here formalizes joint action as a process of interactive inference based on shared task knowledge between the agents [2], [46].

The main contribution of this article is showing that effective cooperative behavior and sensorimotor communication can emerge in dyads of active inference agents that jointly optimize their beliefs about the joint goal and their plans about how to achieve it. We exemplify the functioning of the model in a "joint maze" task. In the task, two agents have to navigate in a maze, to reach and press together either a red or a blue button. Each agent has probabilistic beliefs about the joint task that the dyad is performing, which covers his own and the other agent's contributions (e.g., Should we both press a red or a blue button?). Each agent continuously infers what the joint task is, based on his (stronger or weaker) prior belief and the observation of the other agent's movements toward one of the two buttons. Then, he selects an action (red or blue button press), in a way that simultaneously fulfils a *pragmatic* (i.e., utility maximization) and an *epistemic* (i.e., uncertainty minimization) objective. Here, the *pragmatic* objective prioritizes policies that achieve the joint task efficiently (e.g., by following the shortest route to reach the to-be-pressed button). Rather, the *epistemic* objective prioritizes policies that help the other agent inferring what the joint goal is (e.g., by selecting a longer route that the other agent can easily associate with the goal of pressing the red button).

The next sections are organized as follows. First, we introduce the consensus problem (called "joint maze") we will use throughout this article to explain and validate our approach. Next, we illustrate the main tenets of the interactive inference model of joint action. Then, we present two simulations that illustrate the functioning of the *inter*active inference model. The first simulation shows that over time, the *inter*active inference aligns the joint task representations of the two agents and their behavior, as observed empirically in several joint action studies [18], [24], [47], [48], [49], [50]. In turn, this form of "interactive alignment" (or "generalized synchrony") optimizes the performance of the dyad. The second simulation shows that when agents have asymmetric information about the joint task, the more knowledgeable agent (or "leader") systematically modifies his behavior, to reduce the uncertainty of the less knowledgeable agent (or "follower"), as observed empirically in studies of sensorimotor communication [16], [18]. This *social epistemic action* ensures the success of joint actions despite incomplete information. Finally, we discuss how our model of interactive inference could help better understand various facets of ("leaderless" and "leader–follower") human joint actions, by providing a coherent formal explanation of their dynamics at both brain and behavioral levels.

II. Problem Formulation and Scenario

To illustrate the mechanisms of the *inter*active inference model, we focus on the consensus problem called "joint maze" task, which closely mimics the setting used in a previous human joint action study [34], see Fig. 1. In this task, two agents (represented as a gray hand and a white hand) have to "navigate" in a grid-like maze, reach the location in which the red or blue button is located, and then press it together. The task is completed successfully when both agents "press" the same button, whatever its color (unless stated otherwise).

At the beginning of each simulation, each agent is equipped with some prior knowledge (or preference) about the goal of the task. This prior knowledge is represented as a probabilistic belief, i.e., a probability distribution over four possible task states; these are "both agents will press red," "both agents will press blue," "the white agent will press red and the gray agent will press blue" and "the white agent will press blue and the gray agent will press red." Importantly, in different simulations, the prior knowledge of the two agents can be congruent (if both assign the highest probability to the same state) or incongruent (if they assign the highest probability to different states); certain (if the probability mass is peaked in one state) or uncertain (if the probability mass is spread across all the states). This creates a variety of coordination problems, which span from easy (e.g., if the beliefs of the two agents are congruent and certain) to difficult problems (e.g., if the beliefs are incongruent or uncertain). Each simulation includes several trials, during which each agent follows a perception-action cycle. First, the agent receives an observation about his own position and the position of the other agent. Then, he updates his knowledge about the goal of the task (*task goal inference*) and forms a plan about how to reach it (*plan inference*). Finally, he makes one movement in the maze (by sampling it probabilistically from the plan that he formed). Then, a new perception-action cycle starts.

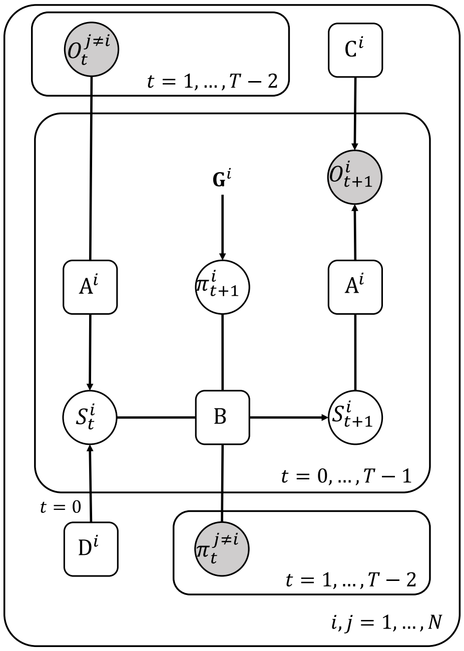

Fig. 2. Generative model for multi-agent active inference. The circles denote stochastic variables. The filled circles denote observed variables and the unfilled circles denote hidden variables that are not observed and need to be inferred. The squares indicate the categorical probability distributions that parameterize the generative model. See the main text for an explanation of the symbols A, B, C, D, G, S, and π. The plate notation indicates that the structures held in a box are replicated, as a function of the indexes. Each agent is denoted with an index of the outer box (from 1 to N) and executes the process in the inner box, whose time horizon is $t = 0, \dots, T-1$. During execution, each agent sends his position and last action and receives the positions and the actions of the other agents. Each agent uses the information from the other agents to infer the evolution of the entire scenario and to update its model. Note that this scheme could be extended to multiple agents.

III. Methods

Here, we provide a brief introduction to the active inference framework for single agents (see [45] for details) and then we illustrate the novel, *inter*active inference model developed here to address multi-agent, cooperative joint actions.

A. Active Inference

Active Inference agents implement perception and action planning through the minimization of variational free energy [45]. To minimize free energy, the agents use a generative model of the causes of their perceptions, which encodes the joint probability of the stochastic variables illustrated in Fig. 2, using the formalism of probabilistic graphical models [51]. The agent's generative model is defined as follows:

$$P(o_{0:T}, s_{0:T}, u_{1:T}, \gamma|\mathbf{\Theta}) = P(\gamma|\mathbf{\Theta})P(\pi|\gamma, \mathbf{\Theta})P(s_0|\mathbf{\Theta})\prod_{t=0}^{T} P(o_t|s_t, \mathbf{\Theta})P(s_{t+1}|s_t, \pi_t, \mathbf{\Theta}) \quad (1)$$

where $P(o_t|s_t, \mathbf{\Theta}) = \mathbf{A}$, $P(s_{t+1}|s_t, \pi_t, \mathbf{\Theta}) = \mathbf{B}(u_t = \pi_t)$, $P(\pi_t|\gamma, \mathbf{\Theta}) = \sigma(\ln \mathbf{E} - \gamma \cdot \mathbf{G}(\pi_t)|\mathbf{\Theta})$, $P(\gamma, \mathbf{\Theta}) \sim \Gamma(\alpha, \beta)$, and $P(s_0|\mathbf{\Theta}) = \mathbf{D}$.

The set $\mathbf{\Theta} = \{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \alpha, \beta\}$ parameterizes the generative model. The (likelihood) matrix $\mathbf{A}$ encodes the relations between the observations O and the hidden causes of observations S. The (transition) matrix $\mathbf{B}$ defines how hidden states evolve over time t, as a function of a control state (action) u_t; note that a sequence of control states $u_1, u_2, \dots, u_t, \dots$ defines a policy π_t (see below for a definition). The matrix $\mathbf{C}$ is an a-priori probability distribution over observations and encodes the agent's preferences or goals. The matrix $\mathbf{D}$ is the prior belief about the initial hidden state, before the agent receives any observation. Finally, $\gamma \in \mathbb{R}$ is a *precision* that regulates action selection and is sampled from a Γ distribution with parameters α and β.

An active inference agent implements the perception-action loop by applying the above matrices to hidden states and observations. In this perspective, perception corresponds to estimating hidden states on the basis of observations and of previous hidden states. At the beginning of the simulation, the agent has access through $\mathbf{D}$ to an initial state estimate S_0 and receives an observation O_0 that permits refining the estimate by using the likelihood matrix $\mathbf{D}$. Then, for $t = 1, \dots, T$, the agent infers its current hidden state S_t based on the observations previously collected and by considering the transitions determined by the control state u_t, as specified in $\mathbf{B}$. Specifically, active inference uses an approximate posterior over (past, present and future) hidden states and parameters $(s_{0:T}, u_{1:T}, \gamma)$. Using the mean field approximation [52], [53], namely, assuming that all variables are independent, the approximate posterior can be factorized and described as

$$Q(s_{0:T}, u_{1:T}, \gamma) = Q(\pi)Q(\gamma)\prod_{t=0}^{T} Q(s_t|\pi_t) \quad (2)$$

where the sufficient statistics are encoded by the expectations $\boldsymbol{\mu} = (\tilde{\mathbf{s}}^{\boldsymbol{\pi}}, \boldsymbol{\pi}, \boldsymbol{\gamma})$, with $\tilde{\mathbf{s}}^{\boldsymbol{\pi}} = \tilde{\mathbf{s}}_0^{\boldsymbol{\pi}}, \dots, \tilde{\mathbf{s}}_T^{\boldsymbol{\pi}}$. Following a variational approach, the distribution in (2) best approximates the posterior when its sufficient statistics μ minimize the free energy of the generative model, see [45]. This condition holds when the sufficient statistics are

$$\mathbf{s}_t^{\boldsymbol{\pi}} \approx \sigma\Big(\ln \mathbf{A} \cdot o_t + \ln\Big(\mathbf{B}(\pi_{t-1}) \cdot \mathbf{s}_{t-1}^{\boldsymbol{\pi}}\Big)\Big) \quad (3a)$$

$$\boldsymbol{\pi} = \sigma(\ln \mathbf{E} - \boldsymbol{\gamma} \cdot \mathbf{G}(\pi_t)) \quad (3b)$$

$$\boldsymbol{\gamma} = \frac{\alpha}{\beta - \mathbf{G}(\boldsymbol{\pi})} \quad (3c)$$

where the symbol "$\cdot$" denotes the inner product, defined as $\mathbf{A} \cdot \mathbf{B} = \mathbf{A}^T\mathbf{B}$, with the two arbitrary matrices $\mathbf{A}$ and $\mathbf{B}$. Action selection is operated by selecting the policy (i.e., sequence of control states $u_1, u_2, \dots, u_t$) that is expected to minimize free energy the most in the future. The policy distribution $\boldsymbol{\pi}$ is expressed in (3b); the term $\sigma(\cdot)$ is a softmax function, $\mathbf{E}$ encodes a prior over the policies (reflecting habitual components of action selection), $\mathbf{G}$ is the expected free energy (EFE) of the policies (reflecting goal-directed components of action selection) and γ is a precision term that encodes the confidence of beliefs about $\mathbf{G}$. The EFE $\mathbf{G}(\pi_t)$ of each policy π_t is defined as

$$\mathbf{G}(\pi_t) = \sum_{\tau=t+1}^{T} D_{KL}[Q(o_\tau|\pi)\|P(o_\tau)] + \mathbb{E}_{\tilde{Q}}[H[P(o_\tau|s_\tau)]] \quad (4)$$

where $D_{KL}[\,\cdot\,\|\,\cdot\,]$ and $H[\,\cdot\,]$ are, respectively, the Kullback–Leibler divergence and the Shannon entropy, $Q(o_\tau, s_\tau|\pi) \triangleq P(o_\tau, s_\tau)Q(s_\tau|\pi)$ is the predicted posterior distribution, $Q(o_\tau|\pi) = \sum_{s_\tau} Q(o_\tau, s_\tau|\pi)$ is the predicted outcome, $P(o_\tau)$ is a categorical distribution representing the preferred outcome and encoded by $\mathbf{C}$, and $P(o_\tau|s_\tau)$ is the likelihood of the generative model encoded by the matrix $\mathbf{A}$.

The EFE can be used as a quality score for the policies and has two terms. The first term of (4) is the Kullback–Leibler divergence between the (approximate) posterior and prior over the outcomes and it constitutes the pragmatic (or utility-maximizing) component of the quality score. This term favors the policies that entail low risk and minimize the difference between predicted ($Q(o_\tau|\pi)$) and preferred ($P(o_\tau) \equiv \mathbf{C}$) future outcomes. The second term of (4) is the expected entropy under the posterior over hidden states and it represents the epistemic component of the quality score. This term favors policies that lead to states that diminish the uncertainty future outcomes $H[P(o_\tau|s_\tau)]$.

After scoring all the policies using EFE, action selection is performed by drawing over the action posterior expectations derived from the sufficient statistic $\boldsymbol{\pi}$ computed via (3b). Then, the selected action is executed, the agent receives a novel observation and the perception-action cycle starts again. See [45] for more details.

B. Multi-Agent Active Inference

The key advancement of this article is extending the active inference framework to a multi-agent setting [1], in which multiple agents (here, two) perform a joint task consisting in navigating in a "joint maze" (Fig. 1) to simultaneously reach either the red or the blue location. The "joint maze" of Fig. 1 includes 21 locations L1, L2, ..., L21. Two agents, gray (i) and white (j), start from the locations L3 and L19 and their goal is to reach either the red (L10) or the blue (L12) goal locations simultaneously. While the two agents can be initially uncertain about their joint goal, to be successful they have to infer what the joint task is and what the best policy or plan is. We refer to these two inferences as "Task goal inference" and "Plan inference," respectively, and we will discuss how they can be realized updating beliefs both within trials (by observing the other's actions) and across trials (by observing whether they were successful).

For each trial, each agent can choose between 25 action sequences or policies π (see Fig. S1 in the supplementary material for their full list), which can be divided into two main types: 1) those that follow the shorter paths that pass through the central corridor or 1) longer paths that go through the maze perimeter. The shorter paths of the gray agent to reach the red and blue goal locations are (L3, L7, L11, L10) and (L3, L7, L11, L12), respectively. The longer paths of the gray agent to reach the red and blue locations are (L3, L2, L1, L6, L9, L10) and (L3, L4, L5, L8, L13, L12), respectively. The shorter paths of the white agent to reach the red and blue goal locations are (L19, L15, L11, L10) and (L19, L15, L11, L12), respectively. The longer paths of the gray agent to reach the red and blue locations are (L19, L18, L17, L14, L9, L10) and (L19, L20, L21, L16, L13, L12), respectively. Below, we will call the first type of policies that go through the shorter paths "pragmatic policies" and the second type of policies that go through the longer paths "social epistemic policies."

Each agent has a separate generative model, whose structure is shown in Fig. 2. In simulation 1, the two agents are equipped with identical generative models, except for a different estimate of their starting locations, L3 or L19. In simulation 2, there are some differences in the $\mathbf{A}$ and $\mathbf{D}$ tensors of the two agents (see below), reflecting the fact that the white agent (the "leader") knows the joint task to be performed, whereas the gray agent (the "follower") does not.

When the two generative models are considered together, they can be defined as $\langle S^i, O^i, U^i, \boldsymbol{\Theta}^i \rangle$, with $i = 1, \ldots, N$, where N is the number of agents (see Fig. 2). Here, we assume that $N = 2$, but it is possible to generalize the model to a larger number of agents. The hidden states $S^i = S^i_1 \otimes S^i_2 \otimes S^i_3$ are obtained as a tensorial product of three vectors (note that unlike the usual algebraic notation for tensors, the subscripts and superscripts do not indicate covariance or contravariance). The three vectors are: the location of the agent S^i_1, the location of the other agent S^i_2, and the agent's belief about the joint goal context S^i_3, or the goals that the agents can jointly achieve. Since there are two potential goals, the agent has a belief about four possible joint goal contexts: 1) blue-blue; 2) blue-red; 3) red-blue; and 4) red-red, encoded as one-hot vectors [1, 0, 0, 0], [0, 1, 0, 0], [0, 0, 1, 0] and [0, 0, 0, 1], respectively. Note that the joint goal context cannot be unambiguously inferred from a single observation and hence inferring it is the most challenging part of the task. In sum, the size of the hidden state is $21 \times 21 \times 4 = 1764$.

The observations $O^i = O^i_1 \otimes O^i_2 \otimes O^i_3 \otimes O^i_4$ consist of the tensorial product between the observed agents' positions O^i_1, O^i_2, the joint goal O^i_3, and the associated utilities O^i_4 (see Fig. S1 in the supplementary material). The first three vectors encode the observations that correspond to the three sets of hidden states S^i_1, S^i_2 and S^i_3. Note that O^i_1 and O^i_2 correspond to the real locations of the two agents; O^i_3 corresponds to a cue about the joint goal that is sampled from A^i_3 (see below); and O^i_4 has always an uninformative value, except at the last time step of the trial, where it corresponds to either a "win" or a "lose." The control states $U^i = U^i_1 \otimes U^i_2$ denote the joint actions available to the agents. Note that in this simulation, each agent has beliefs about his own and the other agent's control states, even if he can only execute his own actions.

The set of tensors $\boldsymbol{\Theta}^i = \{\mathbf{A}^i, \mathbf{B}^i, \mathbf{C}^i, \mathbf{D}^i\}$ defines the structure of the generative model (see Fig. 2). The tensors $\mathbf{D}^i = \mathrm{D}^i_1 \otimes \mathrm{D}^i_2 \otimes \mathrm{D}^i_3$ and $\mathbf{C}^i = \mathrm{C}^i_1 \otimes \mathrm{C}^i_2 \otimes \mathrm{C}^i_3 \otimes \mathrm{C}^i_4$ encode the priors of the hidden states and the observations, respectively. The former factor reflects the agent's prior knowledge about its initial state. We assume that each agent knows his own and the other agent's initial locations (D^i_1 and D^i_2 are deterministic), but the belief D^i_3 about which goal the other expects is uncertain and adjustable as a simulation parameter. Besides, D^i_3 depends on the agent's role, "leader" or "follower." The leader knows the joint task goal and that the follower is uncertain about it; to reflect this, the leader splits the probability mass of D^i_3 equally between blue-blue and blue-red (if he knows that the goal is blue) or between red-red and red-blue (if he knows that the goal is red). Conversely, the follower only knows that, in order to succeed, both agents have to achieve the same goal. Hence, he splits the probability mass of $\mathrm{D}_3^{\mathrm{i}}$ equally between blue-blue and red-red. The factor $\mathrm{C}_4^{\mathrm{i}}$ (i.e., a prior over observations that incentivizes preferred outcomes) depends on the agent's role, in the same way as $\mathrm{D}_3^{\mathrm{i}}$. The (likelihood) mapping between S^i and O^i is specified through the tensor $\mathbf{A}^i$, defined as the tensorial product $\mathbf{A}^i = \bigotimes_k \mathrm{A}_k^i$, $k = 1, \ldots, 4$, where each A_k^i, one for each different factor of O^i, is a 4-order tensor defined on the hidden states.

The first factor $\mathrm{A}_1^{\mathrm{i}}$ is an identity tensor that maps the hidden states that represent the agent's positions in the maze into their corresponding observations. Before defining the tensors $\mathrm{A}_2^{\mathrm{i}}$ and $\mathrm{A}_3^{\mathrm{i}}$ we introduce a *salience* function, which scores the evidence that an agent is pursuing a certain goal, given that he is in a given location.

The salience of the location s_1 with respect to the blue goal is calculated as follows:

$$v_b^{\mathrm{D^i}}(s_1) = \left(\frac{d(s_1, \mathrm{L10})}{d(s_1, \mathrm{L10}) + d(s_1, \mathrm{L12})}\right) \cdot \left(1 - \left(\max\left(\mathrm{D}_3^{\mathrm{i}}\right) - 0.5\right)\right). \quad (5)$$

Equation (5) depends on two terms. The first term implies that the smaller the Euclidean distance between the agent location and the blue goal location (L12 in Fig. 1), the greater the evidence that the agent is pursuing the blue goal (note that we could have used a more complex measure that also considers, for example, direction of movement, but we found that Euclidean distance is sufficient in this setting). The second term implies that the more peaked the mode of $\mathrm{D}_3^{\mathrm{i}}$, the smaller the salience. It could be interpreted as a simple form of attention modulation or precision control [45], [54], which prioritizes bottom-up observations if prior beliefs have low precision. Note that it is possible to define the salience $v_r^{\mathrm{D^i}}(s_1)$ of the state s_1 with respect to the red goal by swapping L10 and L12 in (5).

The tensor $\mathrm{A}_2^{\mathrm{i}}$ maps the hidden states of the agent into the observations O_2 that are relative to the other agent's location. We calculate $\mathrm{A}_2^{\mathrm{i}}$ in two steps:

1) We calculate the (absolute) difference between the salience of one's own and the other agent's location, with respect to the blue goal $v_b^{\mathrm{D^i}}(s_k)$ and the red goal $v_r^{\mathrm{D^i}}(s_k)$, as $\Delta v_k^i = |v_b^{\mathrm{D^i}}(s_k) - v_r^{\mathrm{D^i}}(s_k)|$, where s_k is s_1 when considering one's location and s_2 when considering the other's location.
2) We define the likelihood $p(o_2|s_1, s_2) = \mathrm{sig}(\Delta v_1^i \cdot \Delta v_2^i)$, where $\mathrm{sig}(x) = 1/(1 + \sigma \cdot e^{-\rho x})$ is the parametric logistic function. We assume that $p(o_2|s_1, s_2)$ ranges in the interval (0.75, 1), which we obtain by fixing the parameters of the logistic function as $\sigma = 10$, $\rho = 4$.

This can be interpreted as a form of precision control, which implies that when the agents' locations unambiguously reveal their joint task goals, the precision of the observation is high and $p(o_2|s_1, s_2) = 1$, while when the agents' locations provide poor information about their joint task goal, the precision of the observation is low and $p(o_2|s_1, s_2) = 0.75$.

The tensor $\mathrm{A}_3^{\mathrm{i}}$ encodes the likelihood of the joint goal context of the two agents. Given the hidden state $s = s_1 \otimes s_2 \otimes s_3$, by assuming the saliences $v_b^{\mathrm{D^i}}(s_1)$ and $v_b^{\mathrm{D^i}}(s_2)$ as independent, we define

$$\mathrm{A}_{3\ 1,2,3}^{\mathrm{i}} = p(o_3|s_1, s_2, s_3) \equiv v_{s_3}^{\mathrm{D^i}}(s_1, s_2) = v_{s_3}^{\mathrm{D^i}}(s_1) \cdot v_{s_3}^{\mathrm{D^i}}(s_2).$$

Thus, $\mathrm{A}_{3\ 1,2,3}^{\mathrm{i}}$ represents the likelihood of the joint goal o_3, given that the two agents are in the locations s_1 and s_2, respectively. The values of $\mathrm{A}_3^{\mathrm{i}}$ are shown in Fig. S2, in the supplementary material. In the figure, the rows indicate the values of the joint goal context s_3 and the columns indicate the values of the initial belief about the task goal (i.e., $\max(\mathrm{D}_3^{\mathrm{i}})$). Each matrix is a grid of size 21-21, where the rows and the columns represent the locations s_1 and s_2, respectively, the colors of the cells correspond to the values of the joint salience $v_{s_3}^{\mathrm{D^i}}(s_1, s_2)$. The matrices in the first column ($\max(\mathrm{D}_3^{\mathrm{i}}) = 0.5$) encode saliencies that are more peaked around the joint task goals. The matrices shown in the next two columns encode gradually more uniform and lower-valued saliencies—up until only the goal locations are salient. Note that in the control simulation (described in the section on Simulation 1) in which we prevent interactive inference to take place, we replace the $\mathrm{A}_3^{\mathrm{i}}$ tensor shown in Fig. S2, in the supplementary material with a uniform tensor that does not allow inferring the goal of the other agent from its location.

The tensor $\mathrm{A}_4^{\mathrm{i}}$ is responsible for modeling the relationship between hidden states and outcomes, which can be positive, neutral or negative. $\mathrm{A}_4^{\mathrm{i}}$ is a deterministic sparse tensor. For any hidden state s corresponding to a joint position $s_1 \otimes s_2$ that does not include any goal location, $\mathrm{A}_4^{\mathrm{i}}$ gives a "neutral" outcome. The definition of "positive" and "negative" outcomes varies depending on the joint task goal and the (leader or follower) roles of the agents. In Simulation 1, where both agents are treated as "followers," the outcome is positive if both the agents are in the same goal location (e.g., both are in L10 or both are in L12). Rather, the outcome is negative if only one agent is on a goal location or if the two agents are in two distinct goal locations. In Simulation 2, one of the two agents (the "follower") has the same tensor $\mathrm{A}_4^{\mathrm{i}}$ as in Simulation 1. Rather, the other agent (the "leader") receives a positive outcome if both agents are in the correct goal location (e.g., L10 if the joint task goal is red-red) but a negative outcome if at least one agent is in the incorrect goal location (e.g., L12 if the joint task goal is red-red).

Tensor B encodes a deterministic mapping between hidden states, given the control state u. Note that here the control state u corresponds to a joint action, not to the action of a single agent; hence it is specified as the tensorial product between the vector of the five possible movements of one agent ("up," "down," "left," "right," and "wait") and the vector of the same five movements of the other agent. The tensor B describes how the spatial locations s_1 and s_2 of the agents change as a function of the joint actions $u^i = u_1^i \otimes u_2^i$, such as "up-up," "up-down," "up-left," etc. Note that the transitions regard exclusively the spatial locations. The transitions among joint goal contexts are not modeled explicitly. Rather, the agents have to infer the current joint goal by observing O_3.

The action-perception cycle of the multi-agent active inference model is the same as the single-agent active inference (see Fig. 2), except that the two agents exchange observations between them. Specifically, agent i receives from the agent j the outcome vectors O_1^j and the action u_1^j and vice versa (for simplicity, we allow the agents to send actions to each other; a slightly more complex generative model could have included as inputs *observations* about others' actions rather than directly their actions). Each agent uses this information (plus the observations O_2^i and the actions u_2^i that they computed) to update his beliefs about the hidden states and control states and then to select a course of actions or policy π.

The two key mechanisms of the model are *Task goal inference* and *Plan inference*. Task goal inference corresponds to inferring what the goal of the joint task is, i.e., updating the belief about the four possible task goals (blue-blue, blue-red, red-blue, and red-red). As the task goal is specified at the level of the dyad, in order to infer it, each agent needs to consider both his prior knowledge about the task goal and the movements of the other agent, which are informative about the other agent's task knowledge. Specifically, task goal inference follows a principle of *rational action*; namely, the expectation that the other agent will act efficiently to achieve his goals [55]. Put simply, if an agent observes the other agent moving toward the red (or blue) goal, he updates his joint goal context, by increasing the probability that the goal is red (or blue). Furthermore, both agents update their beliefs about the task goal at the end of each trial, when they receive feedback about success ("win" observation) or failure ("lose" observation).

Plan inference corresponds to inferring the course of action (or plan) that maximizes task success, on the basis of the inferred joint task. In this model, each agent infers both his own and the other agent's plan—although, of course, he can only execute his own plan. The inference about one's own and the other agent's plans needs to consider the utility of following different routes (which privileges the shortest route) and the uncertainty about the goal (which prompts "pistemic" behavior and the selection of informative routes). The balance between utilitarian and epistemic components of planning will become important in Simulation 2, see later.

A key thing to notice is that the perception-action cycles of the two agents—and their inferential processes—are mutually interdependent, as the movements of one agent determine the observations of the other agent at the next time step. Our simulations will show that this *interactive inference* naturally leads to the alignment of belief states and behavioral patterns of the two agents, analogous to the synchronization of neuronal activity and kinematics in socially interacting dyads [15], [30], [49]. Furthermore, the simulations will show that "social epistemic actions" that aim at reducing the uncertainty of the other agent increase the alignment and task success, especially in tasks with asymmetric knowledge.

IV. Simulations of Interactive Inference

We present two simulations of interactive inference using the "joint maze" task of Fig. 1. The first simulation illustrates the case in which both agents know that the goal can be either the blue or the red button and the same goal should be reached by both of them simultaneously. This simulation illustrates that even if the two agents start with uncertain prior on joint goal contexts, during time their beliefs (and behavior) gradually align, which permits the agents to successfully complete the task most of the time. The second simulation illustrates the case of two agents that initially have asymmetric task knowledge. One of them (the leader) knows which one of blue or red is the goal. The other (follower) agent does not know this, but knows that the same goal should be achieved with the leader simultaneously. This simulation illustrates sensorimotor communication—and the importance for the leader to select (epistemic) actions that reduce the follower's uncertainty, in order to complete the task successfully. Note that in these simulations, the agent's beliefs about hidden states $S^i = S_1^i \otimes S_2^i \otimes S_3^i$ change over time, whereas the parameters of the model are assumed to be fixed.

A. Simulation 1 (Leaderless Interaction)

The goal of Simulation 1 is testing whether and how interactive alignment favors the alignment of behavior and belief states of two agents engaged in the "joint maze" task. This simulation comprises 100 trials. For each trial, two identical agents (apart for their prior on the joint goal context, see later) start from two opposite locations of the "joint maze": the gray agent starts in location L3 and the white agent starts in location L19. They can move one step at a time, or wait (i.e., remain in the same location), until they reach one of the locations that include colored goals (red in L10, blue in L12). There are multiple sequences of actions (aka "policies") that each agent can take to reach the goal locations, which correspond to shorter or longer paths, with or without "wait" actions, etc. The 25 policies used in the simulation are specified in Section III. What is most important here is that irrespective of the selected policy, a trial is only successful if both agents reach the same goal / button location, red (L19) or blue (L12). Specifically, if at the end of the trial both agents are in the red (L10) or the blue (L12) button location, then the trial is successful and the agents receive the preferred observation ("win"). Otherwise, if the two agents fail to reach the same button location simultaneously (e.g., one is in L10 and the other is in L12), the trial is unsuccessful and the agents receive an undesirable observation ("lose").

Fig. 3 shows the results of one example simulation. At trial 1, the agents start with the same prior on the joint context goal. This uncertain belief assigns 0.5 to "both agents will press red" (in short, red-red), 0.5 to "both agents will press blue" (in short, blue-blue) and zero to the two other possible states (red-blue: "the white agent will press red and the gray agent will press blue" and blue-red: "the white agent will press blue and the gray agent will press red").

The first two panels of Fig. 3 show the prior beliefs about the joint goal context of the white and gray agents, respectively, at the beginning of each trial, from 1 to 100. In this and the subsequent simulations, the agents' prior beliefs for the first trial are set manually, as discussed above. Then, the prior beliefs are updated within trials, as a result of inference. Furthermore, they are updated across trials: they are the posterior beliefs at the end of the previous trial (as usual in Bayesian inference), but multiplied by a fixed (volatility) factor. This ensures that the prior probability of red-red or blue-blue cannot

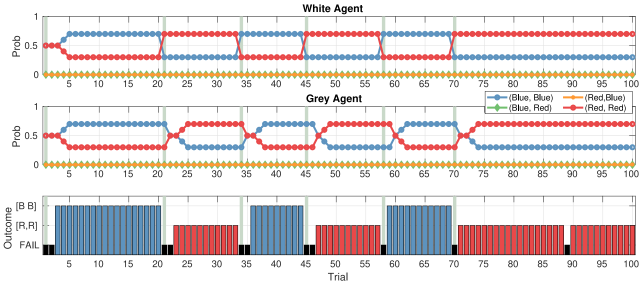

Fig. 3. Results of Simulation 1. The first two panels show the prior beliefs of the white (first panel) and gray (second panel) at the beginning of each trial. The vertical bars indicate moments in which we manually change the mind of the white agent. We *invert* his belief about the joint goal context assigning a higher value (0.7) to blue-blue (if its prior belief assigned higher probability to red-red) or to red-red (if its prior belief assigned higher probability to blue-blue). The third panel shows the outcome of the trials. These include successful trials in which the agents press the blue button (blue bars) or the red button (red bars) and failures (black bars).

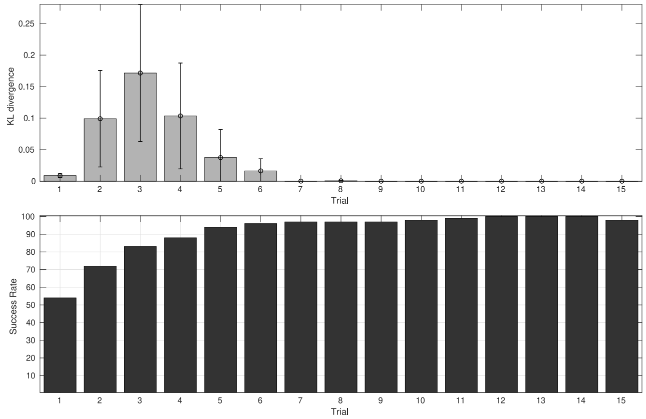

Fig. 4. Average results of 100 runs, with the same parameters as Simulation 1, for 15 trials. The top panel shows a measure of belief (dis)alignment of the agents: the KL divergence between their joint goal contexts. The plot shows the mean value of KL and the standard deviation of the mean (note that we removed outliers whose KL fell outside a confidence interval of 95%). The bottom panel shows a histogram of mean success rate.

be higher than 0.7. This is because in many trials, the posterior beliefs reach the value of 1 for red-red or blue-blue, so the agent is sure about the shared task goal. If this posterior value were used as the prior value for subsequent trials, there would be little place for changes of mind. Introducing the fixed factor amounts to assuming that the agents are not fully sure that joint task goal would remain the same across trials—or in other words, believe that the environment has some volatility. From time to time (vertical bars) we manually "change the mind" of agent 1 from red-red to blue-blue or vice versa, to introduce variability in the simulation.

The third panel of Fig. 3 shows whether the agents completed successfully the trial by pressing the same button (the blue bars indicate that both pressed the blue button, whereas the red bars indicate that both pressed the red button) or unsuccessfully (black bars). Finally, the gray vertical bars show trials in which the white agent "changes mind" about the goal (e.g., from blue-blue to red-red or vice-versa). Following a "change of mind," the dyad usually requires one or a few trials before realigning on the new joint task goal.

Fig. 3 shows that the two agents end up the trials with aligned belief states most of the times, except in the first trials (in which they started with uncertain beliefs) and immediately after the changes of mind (vertical bars). Furthermore, the two agents are successful during most of the trials in which their beliefs are aligned and unsuccessful when their beliefs are not aligned. As shown in Fig. 3, the errors occur in the very first trials, immediately after the gray agent changes mind and in one trial afterward. The errors on the first trials may occur because the agents are uncertain about what to do and they assign the same (EFE) "score" to the two policies that go straight to the red button and the blue button; see Section III for an explanation of EFE and Fig. S3 in the supplementary material for an illustration of the EFE of the policies of the white agent at the beginning of the first trials. When the two agents are very uncertain, there are two possible behaviors:

1) Both agents may select their task goals randomly, which might or might not result in an error (see Fig. S4 in the supplementary material for an illustration of the results of 100 replications of the same experiment, without changes of mind).
2) One agent might simply follow the other and be successful. This "follower effect" is particularly apparent when the agents' prior beliefs are weak, as in the first trials.

Rather, in trials in which the agents' prior beliefs are strong, such as after a change of mind, they do not simply follow one another, but try to fulfill their prior belief—and this explains why we observe several errors after only one of the agents changes mind. These examples illustrate that it is the strength (or the precision) of the beliefs about the joint goal context that determines whether or not an imitative response takes place; see also [41] for a robotic demonstration of the importance of prior beliefs in enabling imitative responses. Finally, note that some errors can occur randomly, with low probability, since action selection is stochastic.

To better quantify the interactive alignment of belief states between the agents across trials and its effects on performance, we executed 100 runs of Simulation 1 and plotted a measure of the belief alignment of the dyad—the KL divergence between the joint goal contexts—and their performance, during the first 15 trials. The top panel of Fig. 4 shows that at the beginning of the simulation, the KL divergence between the prior beliefs of the two agents is small, as they are both equally uncertain about their joint task. While their beliefs are apparently aligned, the alignment regards an uncertain state—and this is why their performance is initially low (see the bottom panel of Fig. 4). During the next few trials, the agents consider different hypotheses about the joint task goal. This leads to a transient increase in the KL divergence between their beliefs (and its variance). After a few trials, the two agents converge to a shared belief about the joint goal and hence the KL divergence decreases. During this process, the success rate increases from 50% (random) to above 90%, within a few trials, see the bottom panel of Fig. 4.

Note that in this simulation the initial choice of a particular joint goal context (red-red or blue-blue) is random, but its persistence across trials depends on a process of interactive belief alignment between the agents. The alignment of behavior and of beliefs about task goals might occur in two ways:

1) It might occur thanks to interactive inference within trials, namely, because each agent monitors the movements of the other agent and uses this information to update his estimate about the joint task goal and the other agent's plan, following a principle of rational action (i.e., the expectation that the other agent will act efficiently to achieve his goals [55]).
2) The alignment might occur because at the end of each trial, the agents receive a feedback about their success ("win" observation) or failure ("lose" observation) and use this feedback to update their beliefs. Second alignment might be the byproduct of a standard reinforcement learning approach to learn behavior by trial and error, without interactive inference within trials (but note that using reinforcement learning would require updating the model parameters, whereas we keep them fixed in our simulations).

To understand whether the first mechanism based on interactive inference is actually useful for alignment and task success, we replicated the same experiment, but by preventing interactive inference to take place. We did this by removing any useful information from the likelihood matrix that maps the others' positions into task goals (i.e., by making the A_3^i tensor uniform; see Section III for details). This control simulation shows that without interactive inference, the performance decreases drastically and there is little alignment: the agents keep switching between red and blue goals and their beliefs do not become increasingly aligned over time (see Figs. S5 and S6 in the supplementary material). This control simulation shows that despite the task could be addressed using (reinforcement-based) feedback from successes and failures, the interactive inference is key to achieve alignment. Increasing the weight assigned to feedback information could potentially increase success rate and alignment, but this does not seem necessary when interactive inference is in place.

In sum, Simulation 1 shows that two agents that engage in interactive inference can align both their joint goal contexts and their plans to achieve the joint task goal, forming shared task knowledge [18], [56], [57]. The alignment at both the belief and behavioral levels is made possible by a process of interactive, reciprocal inference of goals and plans. The two agents initially have weak beliefs about the goal identity and therefore they can "follow each other" until they settle on some joint goal—and successively stick to it.

B. Simulation 2 (Asymmetric Leader–Follower Interaction)

The goal of Simulation 2 is testing the emergence of "leader–follower" dynamics observed in human studies using the "joint maze" setup [34] and other related studies in which the agents have asymmetric preferences (or information) about the joint task goal [25], [58], [59], [60]. This simulation is similar to Simulation 1, but the two agents differ in their prior beliefs about the task goal [25], [34], [58], [59], [60]. Specifically, the white agent (the "leader") knows the task to be accomplished—for example, red-red—whereas the gray agent (the "follower") does not. In other words, while in Simulation 1 both agents had initially weak beliefs (or preferences) about the joint goal and can be therefore considered two "followers," in Simulation 2 one of the two agents is a "leader" and has a strong initial preference about the joint task goal.

The generative model of the follower is identical to the one used in Simulation 1, whereas the generative model of the leader differs from it in two ways. First, the (likelihood) tensor A_4^i of the white agent reflects his knowledge of the true task contingencies; namely, that the preferred "win" observation can only be obtained by achieving the joint task red-red, but not blue-blue (or the opposite when the true task goal is blue-blue). Furthermore, the prior belief of the white agent is 0.5 for red-red, 0.5 for red-blue and 0 for the two other joint goals (or the opposite when the true task goal is blue-blue). This prior belief is updated both within and across trials. as in Simulation 1.

Several studies [25], [34], [58], [59], [60] showed that when leaders and followers have asymmetric information, the leaders modify their movement kinematics to *signal* their intentions and reduce the uncertainty of the followers [16], [35]. For example, consider that in the scenario of Fig. 1 the leader (white agent) has a choice between two kinds of action sequences or policies to reach the red goal location. The first, "pragmatic policies" follow the shortest and hence most efficient path to the goal: L15, L11, and L10. However, if the leader selects the pragmatic policy, he does not offer any cue to the follower about the joint task goal, until the last action (to L10). This is because passing through L15 and L11 is equally likely if the intended goals are red or blue and hence does not provide diagnostic information about the goal location. The second, "social epistemic policies" follow the route through L18, L17, L14, L9 and L10, which, despite being longer, provides to the follower early information to the intended goal location. This is because passing through L18, L17, L14, L9 is rational only if the goal is the red button—and hence it provides diagnostic evidence that the to-be-pressed button is red. The above studies [25], [34], [58], [59], [60] show that leaders often select "social epistemic policies": they sacrifice efficiency to reduce the follower's uncertainty.

The tradeoff between pragmatic and epistemic components of policy selection is automatic in active inference, because the EFE functional used in active inference to score policies includes two components: 1) a "pragmatic component" that maximizes utility and prioritizes the shortest paths to the goal and 2) an "pistemic component" that minimizes uncertainty (see Section III). We therefore expected the leaders to select "social epistemic policies" most often when the followers were uncertain—and select "pragmatic policies" when uncertainty was reduced.

The results of an example leader–follower simulation lasting 30 trials are shown in Fig. 5. The first and third panels of Fig. 5 show the prior beliefs of the leader (white agent) and the follower (gray agent), respectively, at the beginning of each trial. These are largely aligned, except in the very first trials. The second and fourth panels show the policies selected by the leader and the follower, respectively. As discussed above, we

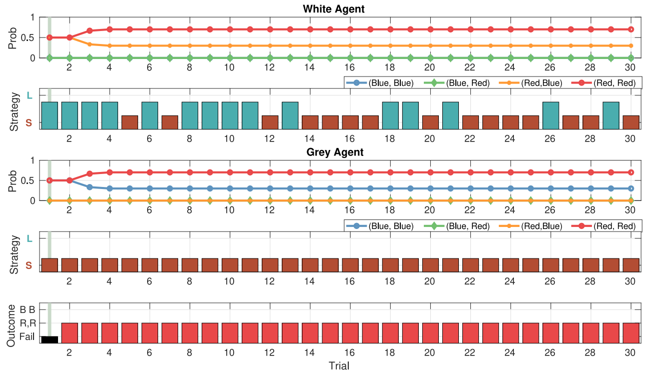

Fig. 5. Results of Simulation 2: example of leader–follower dyadic interaction, for 30 trials. The first two panels show the prior beliefs of the leader at the end of each trial and the policy he selects (shorter red bar: pragmatic policy that follows the shortest path to the goal; longer blue bar: social epistemic policy that follows the longer but more informative path to the goal). The third and fourth panels show the prior beliefs of the follower at the end of each trial and the policy he selects. The fifth panel shows the outcome of the trials. These include successful trials in which the agents press the red button (red bars) and failures (black bars).

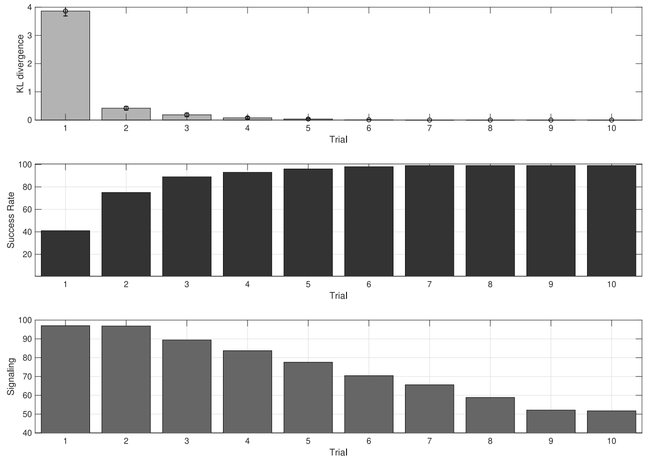

Fig. 6. Average results of 100 runs, with the same parameters as Simulation 2. The format of the first two plots is the same as Fig. 4. The third plot shows the percentage of policies selected by the leader that we label as "social epistemic actions" and prescribe signaling behavior. See the main text for explanation. For example, if the gray agent is the leader, he can select an epistemic policy that passes through L3, L2, L1, L6, L9, and L10 (to reach the red button) or through L3, L4, L5, L8, L13, and L12 (to reach the blue button).

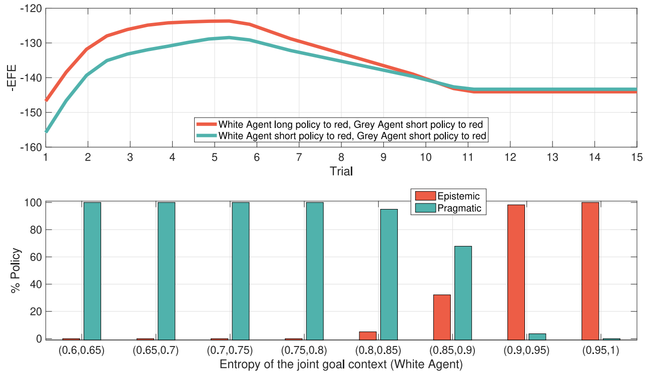

Fig. 7. How the leader balances epistemic and pragmatic policies in Simulation 2. Top panel: negative EFE of the most useful epistemic (red) and pragmatic policy (green), in the first ten trials. Bottom panel: frequency of the most useful epistemic policy as a function of the entropy of the leader's joint goal context. This entropy provides a measure of the (leader's estimate of the) follower's uncertainty.

divided policies into two categories: 1) "pragmatic policies" (S: shorter red bars) that follow the shortest path to the goal and 2) "social epistemic policies" (L: longer blue bars) that follow longer but more informative paths. The second panel of Fig. 5 shows that the leader tends to select "social epistemic policies" in the first trials, to reduce the follower's uncertainty (see also Fig. S7 in the supplementary material for a visualization of the EFE of the leader's policies). Rather, the follower has no benefit from selecting epistemic policies and selects pragmatic policies across almost all the trials. Finally, the bottom panel of Fig. 5 shows that in all but the first two trials (short black bars), the agents successfully achieve the red-red goal (long red bars).

Fig. 6 shows the results of 100 repetitions of the same simulation (see also Fig. S8 in the supplementary material). The first panel shows that the beliefs of the agents, measured as the KL divergence between their prior beliefs about the task goal, align over time. The second panel shows that the average performance of the dyads, measured as the number of times they select the correct red-red goal, increases over time. The third panel shows the percentage of "social epistemic policies" selected by the leaders. Initially, the leaders have a strong tendency to select "social epistemic policies," but this tendency decreases significantly across trials, as the followers become increasingly certain about the joint task goal, as shown empirically [25], [34], [58], [59], [60]. This result emerges because in the EFE (used in active inference to score policies), the decrease of uncertainty lowers the epistemic value of policies, hence lowering the probability that they will be selected [61].

Fig. 7 permits appreciating how the leader balances epistemic and pragmatic policies over time. The first panel shows the negative EFE (averaged across 100 repetitions) that the leader selects the most useful social epistemic policy (red line) and the most useful pragmatic policy (green line). It shows that the social epistemic policy has a very high probability during the first five trials, then its probability decreases until the pragmatic policy becomes the most likely, starting from trial 10. The second panel shows the probability (averaged across 100 repetitions) that the leader selects a social epistemic policy as a function of his uncertainty about the task goal, quantified as the entropy of his belief about the joint goal context (please remind that this is a shared representation that encodes both the leader's and the follower's contributions). The entropy over this variable reflects an estimate of the follower's uncertainty, not of the leader's uncertainty (as the leader knows the goal) and decreases over time, as the follower becomes less uncertain. Notably, the results shown in Fig. 7B in the supplementary material closely correspond to the findings of a study that uses the "joint maze" setting [34]. The study reports that the probability that a (human) participant selects a pragmatic policy is high only when (his or her estimate of) the follower's uncertainty is very low (see [34, Fig. 8A]), which is in good agreement with the pattern shown in our Fig. 7B in the supplementary material.

In sum, Simulation 2 shows that in leader–follower interactions with asymmetric knowledge, leaders select "social epistemic policies"—and therefore sacrifice some efficiency in their choice of movements—to signal their intended goals to the followers and reduce their uncertainty. This signaling behavior is progressively reduced, when the followers become more certain about the joint action goal. This form of signaling was shown in previous computational models that used goal and plan inference, but the models used ad-hoc formulations to promote social epistemic actions [16], [35]. In contrast, social epistemic actions emerge naturally in our model, for two reasons. First, the EFE functional used to score policies balances automatically pragmatic and epistemic components. This means that when uncertainty resolution is necessary, the EFE functional automatically promotes epistemic behavior [45]. To illustrate this point, we performed a control simulation (Fig. S9 in the supplementary material) that is the same as Simulation 2, except that we removed the "epistemic component" from the EFE (see Section III). In the control simulation, the leader selects significantly less social epistemic policies, the behavioral alignment process is slower and the success rate grows more slowly compared to the case in which the EFE is used. This control simulation shows that EFE affords social epistemic actions and these promote leader–follower interactions.

The second reason why social epistemic behavior emerges in our model is because the leader's generative model includes beliefs about the shared task goal. When scoring his policies (via the EFE functional), the leader considers the uncertainty (or the entropy) of the shared task goal and it assigns high probability to "epistemic" policies, regardless of the fact that they lower his own uncertainty (as shown in previous studies) or the follower's uncertainty [34]. This illustrates that active inference agents endowed with shared representations behave in socially oriented ways, even without ad-hoc incentives.

V. Conclusion

Joint actions are ubiquitous in our lives and engage several cognitive abilities, such as mutual prediction, mental state inference, sensorimotor communication, and shared task representation. However, we still lack a comprehensive formal model that explains these abilities from first principles. Here, we proposed a computational model of *inter*active inference, in which two active inference agents coordinate around a joint goal "—pressing together either a red or a blue button"—that they do not know in advance (Simulation 1) or that only one of them knows in advance (Simulation 2).

Our results show that the interactive inference model can successfully reproduce key behavioral and neural signatures of dyadic interactions. Simulation 1 shows that when two agents have the same (uncertain) knowledge about the joint task to be performed, they spontaneously coordinate around a joint goal and align their behavior and their task knowledge (here, their beliefs about the joint goal) over time. This result is in keeping with a large number of studies showing the synchronization of neuronal activity and kinematics during joint actions, perhaps as a way to enhance coordination and the sense of joint agency [15], [30], [49], [62]. Furthermore, interactive inference is robust to sudden changes of mind of one of the agents, as indexed by the fact that the alignment of behavior and task knowledge is recovered fast. While simple joint tasks, such as the "joint maze," that we adopted could be in principle learned by trial and error and without inference, our control simulation shows that interactive inference within trials promotes better performance and alignment of behavior and of belief states (see Figs. S5 and S6 in the supplementary material).

Simulation 2 shows that during dyadic interactions in which only one agent (the "leader") knows the task to be performed but the other agent (the "follower") does not, the leader systematically selects "social epistemic policies" in early trials. The social epistemic policies sacrifice some path efficiency to give the follower early cues about the task goal, hence reducing his/her uncertainty and contributing to optimize the joint action. The results of this simulation are in keeping with a large number of studies of *sensorimotor communication* during dyadic interactions with asymmetric information [25], [34], [58], [59], [60]. Specifically, our model reproduces two key phenomena of leader–follower interactions:

1) In all these tasks, leaders select an apparently less efficient path, which however provides early information about the intended task goal.
2) The selection of these more informative (or social epistemic) policies is dependent on the follower's uncertainty and it is abolished when the follower is (estimated to be) no longer uncertain, as reported in a study that uses our "joint maze" setup [34] and other studies in which the uncertainty of the follower varies across trials [25], [60].

Different from previous models, here the leader's social epistemic behavior does not require ad-hoc mechanisms [16], [35]. Rather, it is a necessary consequence of the fact that the agents have shared task knowledge and select actions using the *EFE* functional, which considers epistemic actions on equal ground with pragmatic actions. In other words, active inference agents who cooperate in uncertain conditions and have beliefs about their shared goal can natively select "epistemic" policies that reduce their own uncertainty (as shown in [45]) and the uncertainty of the other agents (as shown in our simulations).

Another important feature of our model is its flexibility. Simulations 1 and 2 use exactly the same computational model, except for the fact that in Simulation 2 only the "leader" knows the goal. This implies that active inference is flexible enough to reproduce various aspects of joint action dynamics, without ad-hoc changes of the model. In our simulations, the differences between standard, "leaderless" (Simulation 1) and "leader–follower" (Simulation 2) dynamics emerge as an effect of the strength (and the precision) of the agents' beliefs about the joint goal to be performed. When the agents' beliefs are uncertain, as in Simulation 1, they tend to follow each other to optimize the joint goal—and update (and align) their beliefs afterward. In this case, the joint outcome (e.g., red-red or blue-blue) can be initially stochastic, but is successively stabilized thanks to the interactive inference. This setting therefore exemplifies a "peer-to-peer" or a "follower–follower" interaction. Yet, it is possible to observe some "leader–follower" dynamics, in the sense that one of the two agents drives the choice of one particular joint task goal. However, in Simulation 1, the role of the leader is not predefined, but rather emerges during the task, as one of the joint goals is stochastically selected during the interaction—and then the two agents stick to it (note however that the situation is different during changes of mind, because the goal is predefined by us rather than being stochastically selected during the interaction). Rather, Simulation 2 exemplifies the case of a "leader–follower" setup in which the role of the leader is predefined—because the leader has a strong preference for one of the goals. The comparison of Simulations 1 and 2 shows that what defines leaders and followers is simply the strength of the prior about the joint goal context (and of its associated outcomes). Our results in Simulation 2 are in keeping with previous active inference models that showed the emergence of behavior synchronization and leader–follower dynamics in joint singing [63] and robotic dyadic interactions [41]. These studies nicely illustrate that several facets of joint actions emerge when two agents infer each other's goals and plans. However, the results reported here go beyond the above studies, by demonstrating the emergence of sensorimotor communication and social epistemic actions when the agents have asymmetric information.

In sum, our simulations provide a proof-of-concept that *interactive inference* can reproduce key empirical results of joint action studies, such as the interactive alignment and synchronization of behavior and neuronal activity (which, in our model, correspond to the belief dynamics) during standard joint actions [15], [30], [49] and the "sensorimotor communication" during dyadic leader–follower joint actions with asymmetric information [25], [34], [58], [59], [60]. An open objective for future research is extending the empirical validation of this framework by adopting it to model more cases of joint action, beyond the "joint maze" scenario of [34]. Another objective for future research is exploiting this framework to design more effective agents that exploit sensorimotor communication to enhance human–robot joint actions. The ease of human–human collaboration rests on our advanced abilities to infer intentions and plans, align representations and select movements that are easily legible and interpretable by our agents [16]. Endowing robots with similar advanced cognitive abilities would permit them to achieve unprecedented levels of success in human–robot collaboration and plausibly increase the trust in robotic agents [44], [64], [65], [66]. Scaling up the approach from the current grid-world simulation to noisy continuous space robotic experiments would require more effective methods to learn sophisticated generative models (using for example deep learning methods [67], [68], [69]) and to plan in large state spaces (using for example tree search methods [70], [71]). Another challenge consists in scaling up the approach to groups of agents. In principle, each agent could maintain beliefs about each group member. However, a more parsimonious alternative is maintaining beliefs over the "average group member's mind"; this latter approach remains to be evaluated in our setting [72]. Finally, a key challenge for future research is extending this framework beyond cooperative joint actions, to also cover competitive and mixed cooperative-competitive interactions, which are frequent in multi-agent settings [73].

Acknowledgment

The GEFORCE Quadro RTX6000 and Titan GPU cards used for this research were donated by the NVIDIA Corporation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

[1] A. Dorri, S. S. Kanhere, and R. Jurdak, "Multi-agent systems: A survey," *IEEE Access*, vol. 6, pp. 28573–28593, 2018.

[2] J. Ferber and G. Weiss, *Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence*, vol. 1. Reading, MA, USA: Addison-Wesley, 1999.

[3] B. Liu, H. Su, R. Li, D. Sun, and W. Hu, "Switching controllability of discrete-time multi-agent systems with multiple leaders and time-delays," *Appl. Math. Comput.*, vol. 228, pp. 571–588, Feb. 2014.

[4] B. Liu, T. Chu, L. Wang, Z. Zuo, G. Chen, and H. Su, "Controllability of switching networks of multi-agent systems," *Int. J. Robust Nonlinear Control*, vol. 22, no. 6, pp. 630–644, 2012.

[5] S. Su, Z. Lin, and A. Garcia, "Distributed synchronization control of multiagent systems with unknown nonlinearities," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 325–338, Jan. 2016.

[6] B. D. Anderson, B. Fidan, C. Yu, and D. Walle, "UAV formation control: Theory and application," in *Recent Advances in Learning and Control*. London, U.K.: Springer, 2008, pp. 15–33.

[7] N. Krothapalli and A. V. Deshmukh, "Distributed task allocation in multi-agent systems," in *Proc. Inst. Ind. Eng. Annu. Conf.*, 2002, pp. 1–6.

[8] W. Ni and D. Cheng, "Leader-following consensus of multi-agent systems under fixed and switching topologies," *Syst. Control Lett.*, vol. 59, nos. 3–4, pp. 209–217, 2010.

[9] V. Trianni, D. De Simone, A. Reina, and A. Baronchelli, "Emergence of consensus in a multi-robot network: From abstract models to empirical validation," *IEEE Robot. Autom. Lett.*, vol. 1, no. 1, pp. 348–353, Jan. 2016.

[10] W. He, B. Xu, Q.-L. Han, and F. Qian, "Adaptive consensus control of linear multiagent systems with dynamic event-triggered strategies," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 2996–3008, Jul. 2020.

[11] C. Castelfranchi, "Modeling social interaction for ai agents," in *Proc. 15th Int. Joint Conf. Artif. Intell.*, 1997, pp. 1567–1576.

[12] R. Conte and C. Castelfranchi, *Cognitive and Social Action*. London, U.K.: Psychology Press, 1995.

[13] C. Castelfranchi and R. Falcone, *Trust Theory: A Socio-Cognitive and Computational Model*. Chichester, U.K.: Wiley, 2010.

[14] R. Sun, "Cognitive science meets multi-agent systems: A prolegomenon," *Philos. Psychol.*, vol. 14, no. 1, pp. 5–28, 2001.

[15] N. Sebanz, H. Bekkering, and G. Knoblich, "Joint action: Bodies and minds moving together," *Trends Cogn. Sci.*, vol. 10, no. 2, pp. 70–76, 2006.

[16] G. Pezzulo, F. Donnarumma, H. Dindo, A. D'Ausilio, I. Konvalinka, and C. Castelfranchi, "The body talks: Sensorimotor communication and its brain and kinematic signatures," *Phys. Life Rev.*, vol. 28, pp. 1–21, Mar. 2019.

[17] O. Oullier and J. A. S. Kelso, "Social coordination, from the perspective of coordination dynamics," in *Encyclopedia of Complexity and Systems Science*, vol. 19. New York, NY, USA: Springer, 2009, pp. 8198–8213.

[18] G. Knoblich and N. Sebanz, "Evolving intentions for social interaction: From entrainment to joint action," *Philos. Trans. Royal Soc. B, Biol. Sci.*, vol. 363, no. 1499, pp. 2021–2031, 2008.

[19] J. Henrich and M. Muthukrishna, "The origins and psychology of human cooperation," *Annu. Rev. Psychol.*, vol. 72, pp. 207–240, Jan. 2021.

[20] G. Pezzulo, F. Donnarumma, S. Ferrari-Toniolo, P. Cisek, and A. Battaglia-Mayer, "Shared population-level dynamics in monkey premotor cortex during solo action, joint action and action observation," *Progr. Neurobiol.*, vol. 210, Mar. 2022, Art. no. 102214.

[21] G. Pezzulo, P. Iodice, F. Donnarumma, H. Dindo, and G. Knoblich, "Avoiding accidents at the champagne reception: A study of joint lifting and balancing," *Psychol. Sci.*, vol. 28, no. 3, pp. 338–345, 2017.

[22] G. Pezzulo, L. Roche, and L. Saint-Bauzel, "Haptic communication optimises joint decisions and affords implicit confidence sharing," *Sci. Rep.*, vol. 11, no. 1, p. 1051, 2021.

[23] G. Knoblich and J. S. Jordan, "Action coordination in groups and individuals: Learning anticipatory control," *J. Exp. Psychol. Learn., Memory, Cogn.*, vol. 29, no. 5, p. 1006, 2003.

[24] R. P. van der Wel, C. Becchio, A. Curioni, and T. Wolf, "Understanding joint action: Current theoretical and empirical approaches," *Acta Psychologica*, vol. 215, Apr. 2021, Art. no. 103285.

[25] M. Candidi, A. Curioni, F. Donnarumma, L. M. Sacheli, and G. Pezzulo, "Interactional leader–follower sensorimotor communication strategies during repetitive joint actions," *J. Royal Soc. Interface*, vol. 12, no. 110, p. 644, 2015.

[26] F. Visco-Comandini et al., "Do non-human primates cooperate? Evidences of motor coordination during a joint action task in macaque monkeys," *Cortex*, vol. 70, pp. 115–127, Sep. 2015.

[27] C. Becchio, A. Koul, C. Ansuini, C. Bertone, and A. Cavallo, "Seeing mental states: An experimental strategy for measuring the observability of other minds," *Phys. Life Rev.*, vol. 24, pp. 67–80, Mar. 2018.

[28] M. I. Coco et al., "Multilevel behavioral synchronization in a joint tower-building task," *IEEE Trans. Cogn. Devel. Syst.*, vol. 9, no. 3, pp. 223–233, Sep. 2017.

[29] A. D'Ausilio, G. Novembre, L. Fadiga, and P. E. Keller, "What can music tell us about social interaction?" *Trends Cogn. Sci.*, vol. 19, no. 3, pp. 111–114, 2015.

[30] P. E. Keller, G. Novembre, and M. J. Hove, "Rhythm in joint action: Psychological and neurophysiological mechanisms for real-time interpersonal coordination," *Philos. Trans. Royal Soc. B, Biol. Sci.*, vol. 369, no. 1658, 2014, Art. no. 20130394.

[31] D. M. Wolpert, K. Doya, and M. Kawato, "A unifying computational framework for motor control and social interaction," *Philos. Trans. Royal Soc. London Series B, Biol. Sci.*, vol. 358, no. 1431, pp. 593–602, 2003.

[32] H. Dindo, D. Zambuto, and G. Pezzulo, "Motor simulation via coupled internal models using sequential Monte Carlo," in *Proc. 22nd Int. Joint Conf. Artif. Intell.*, 2011, pp. 2113–2119.

[33] H. Dindo, F. Donnarumma, F. Chersi, and G. Pezzulo, "The intentional stance as structure learning: A computational perspective on mindreading," *Biol. Cybern.*, vol. 109, pp. 453–467, Jul. 2015.

[34] G. Pezzulo and H. Dindo, "What should I do next? Using shared representations to solve interaction problems," *Exp. Brain Res.*, vol. 211, nos. 3–4, pp. 613–630, 2011.

[35] G. Pezzulo, F. Donnarumma, and H. Dindo, "Human sensorimotor communication: A theory of signaling in online social interactions," *PloS One*, vol. 8, no. 11, 2013, Art. no. e79876.

[36] C. L. Baker, R. Saxe, and J. B. Tenenbaum, "Action understanding as inverse planning," *Cognition*, vol. 113, no. 3, pp. 329–349, 2009.

[37] T. Ullman, C. Baker, O. Macindoe, O. Evans, N. Goodman, and J. Tenenbaum, "Help or hinder: Bayesian models of social goal inference," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 22, 2009, pp. 1–9.

[38] N. Tang, S. Stacy, M. Zhao, G. Marquez, and T. Gao, "Bootstrapping an imagined we for cooperation," in *Proc. CogSci*, 2020, pp. 1–6.

[39] J. Hwang, J. Kim, A. Ahmadi, M. Choi, and J. Tani, "Dealing with large-scale spatio-temporal patterns in imitative interaction between a robot and a human by using the predictive coding framework," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 5, pp. 1918–1931, May 2020.

[40] J. Tani, R. Nishimoto, J. Namikawa, and M. Ito, "Codevelopmental learning between human and humanoid robot using a dynamic neural-network model," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 1, pp. 43–59, Feb. 2008.

[41] N. Wirkuttis and J. Tani, "Leading or following? Dyadic robot imitative interaction using the active inference framework," *IEEE Robot. Autom. Lett.*, vol. 6, no. 3, pp. 6024–6031, Jul. 2021.

[42] K. Dautenhahn et al., "How may I serve you? A robot companion approaching a seated person in a helping context," in *Proc. 1st ACM SIGCHI/SIGART Conf. Human-Robot Interact.*, 2006, pp. 172–179.

[43] A. Clodic and R. Alami, "What is it to implement a human–robot joint action?" in *Robotics, AI, and Humanity: Science, Ethics, and Policy*. Cham, Switzerland: Springer, 2021, pp. 229–238.

[44] A. Clodic, E. Pacherie, R. Alami, and R. Chatila, "Key elements for human–robot joint action," in *Sociality and Normativity for Robots: Philosophical Inquiries into Human-Robot Interactions*. Cham, Switzerland: Springer, 2017, pp. 159–177.

[45] T. Parr, G. Pezzulo, and K. J. Friston, *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. Cambridge, MA, USA: MIT Press, 2022.

[46] Q. Song, F. Liu, H. Su, and A. V. Vasilakos, "Semi-global and global containment control of multi-agent systems with second-order dynamics and input saturation," *Int. J. Robust Nonlinear Control*, vol. 26, no. 16, pp. 3460–3480, 2016.

[47] I. Konvalinka, P. Vuust, A. Roepstorff, and C. D. Frith, "Follow you, follow me: Continuous mutual prediction and adaptation in joint tapping," *Quart. J. Exp. Psychol.*, vol. 63, no. 11, pp. 2220–2230, 2010.

[48] J. C. Skewes, L. Skewes, J. Michael, and I. Konvalinka, "Synchronised and complementary coordination mechanisms in an asymmetric joint aiming task," *Exp. Brain Res.*, vol. 233, pp. 551–565, Feb. 2015.

[49] G. Novembre, G. Knoblich, L. Dunne, and P. E. Keller, "Interpersonal synchrony enhanced through 20 Hz phase-coupled dual brain stimulation," *Social Cogn. Affect. Neurosci.*, vol. 12, no. 4, pp. 662–670, 2017.

[50] S. Garrod and M. J. Pickering, "Joint action, interactive alignment, and dialog," *Topics Cogn. Sci.*, vol. 1, no. 2, pp. 292–304, 2009.

[51] C. M. Bishop and N. M. Nasrabadi, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[52] C. Zhang, J. Bütepage, H. Kjellström, and S. Mandt, "Advances in variational inference," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 2008–2026, Aug. 2019.

[53] M. Opper and D. Saad, *Advanced Mean Field Methods: Theory and Practice*. Cambridge, MA, USA: MIT Press, 2001.

[54] D. Maisto, F. Donnarumma, and G. Pezzulo, "Nonparametric problem-space clustering: Learning efficient codes for cognitive control tasks," *Entropy*, vol. 18, no. 2, p. 61, 2016.

[55] G. Gergely and G. Csibra, "Teleological reasoning in infancy: The Naıve theory of rational action," *Trends Cogn. Sci.*, vol. 7, no. 7, pp. 287–292, 2003.

[56] N. Sebanz, G. Knoblich, W. Prinz, and E. Wascher, "Twin peaks: An ERP study of action planning and control in coacting individuals," *J. Cogn. Neurosci.*, vol. 18, no. 5, pp. 859–870, 2006.

[57] N. Sebanz, G. Knoblich, and W. Prinz, "How two share a task: Corepresenting stimulus-response mappings," *J. Exp. Psychol. Human Percept. Perform.*, vol. 31, no. 6, p. 1234, 2005.

[58] L. M. Sacheli, E. Tidoni, E. F. Pavone, S. M. Aglioti, and M. Candidi, "Kinematics fingerprints of leader and follower role-taking during cooperative joint actions," *Exp. Brain Res.*, vol. 226, pp. 473–486, May 2013.

[59] C. Vesper and M. J. Richardson, "Strategic communication and behavioral coupling in asymmetric joint action," *Exp. Brain Res.*, vol. 232, pp. 2945–2956, May 2014.

[60] F. Leibfried, J. Grau-Moya, and D. A. Braun, "Signaling equilibria in sensorimotor interactions," *Cognition*, vol. 141, pp. 73–86, Aug. 2015.

[61] K. Friston, T. FitzGerald, F. Rigoli, P. Schwartenbeck, and G. Pezzulo, "Active inference: A process theory," *Neural Comput.*, vol. 29, no. 1, pp. 1–49, 2017.

[62] M. Shiraishi and S. Shimada, "Inter-brain synchronization during a cooperative task reflects the sense of joint agency," *Neuropsychologia*, vol. 154, Apr. 2021, Art. no. 107770.

[63] K. Friston and C. Frith, "A duet for one," *Consciousness Cogn.*, vol. 36, pp. 390–405, Nov. 2015.

[64] K. Belhassein et al., "Addressing joint action challenges in HRI: Insights from psychology and philosophy," *Acta Psychologica*, vol. 222, Feb. 2022, Art. no. 103476.

[65] F. Donnarumma, H. Dindo, and G. Pezzulo, "Sensorimotor communication for humans and robots: Improving interactive skills by sending coordination signals," *IEEE Trans. Cogn. Devel. Syst.*, vol. 10, no. 4, pp. 903–917, Dec. 2018.

[66] A. D. Dragan, K. C. Lee, and S. S. Srinivasa, "Legibility and predictability of robot motion," in *Proc. 8th ACM/IEEE Int. Conf. Human-Robot Interact. (HRI)*, 2013, pp. 301–308.

[67] K. Ueltzhöffer, "Deep active inference," *Biol. Cybern.*, vol. 112, no. 6, pp. 547–573, 2018.

[68] T. Taniguchi et al., "World models and predictive coding for cognitive and developmental robotics: Frontiers and challenges," 2023, *arXiv:2301.05832*.

[69] P. Lanillos et al., "Active inference in robotics and artificial agents: Survey and challenges," 2021, *arXiv:2112.01871*.

[70] D. Maisto, F. Gregoretti, K. Friston, and G. Pezzulo, "Active tree search in large POMDPs," 2021, *arXiv:2103.13860*.

[71] G. Pezzulo, F. Donnarumma, D. Maisto, and I. Stoianov, "Planning at decision time and in the background during spatial navigation," *Current Opin. Behav. Sci.*, vol. 29, pp. 69–76, Oct. 2019.

[72] K. Khalvati et al., "Modeling other minds: Bayesian inference explains human choices in group decision-making," *Sci. Adv.*, vol. 5, no. 11, 2019, Art. no. eaax8783.

[73] T. T. Nguyen, N. D. Nguyen, and S. Nahavandi, "Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications," *IEEE Trans. Cybern.*, vol. 50, no. 9, pp. 3826–3839, Sep. 2020.

Open Access funding provided by 'Consiglio Nazionale delle Ricerche-CARI-CARE-ITALY' within the CRUI CARE Agreement