frontiers in Psychology

ORIGINAL RESEARCH
published: 23 February 2017
doi: 10.3389/fpsyg.2017.00237

Sensorimotor Coarticulation in the Execution and Recognition of Intentional Actions

Francesco Donnarumma[1], *Haris Dindo*[2] *and Giovanni Pezzulo*[1]*

[1] *Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy,* [2] *Computer Science Engineering, University of Palermo, Palermo, Italy*

Humans excel at recognizing (or inferring) another's distal intentions, and recent experiments suggest that this may be possible using only subtle kinematic cues elicited during early phases of movement. Still, the cognitive and computational mechanisms underlying the recognition of intentional (sequential) actions are incompletely known and it is unclear whether kinematic cues alone are sufficient for this task, or if it instead requires additional mechanisms (e.g., prior information) that may be more difficult to fully characterize in empirical studies. Here we present a computationally-guided analysis of the execution and recognition of intentional actions that is rooted in theories of motor control and the coarticulation of sequential actions. In our simulations, when a performer agent coarticulates two successive actions in an action sequence (e.g., "reach-to-grasp" a bottle and "grasp-to-pour"), he automatically produces kinematic cues that an observer agent can reliably use to recognize the performer's intention early on, during the execution of the first part of the sequence. This analysis lends computational-level support for the idea that kinematic cues may be sufficiently informative for early intention recognition. Furthermore, it suggests that the social benefits of coarticulation may be a byproduct of a fundamental imperative to optimize sequential actions. Finally, we discuss possible ways a performer agent may combine automatic (coarticulation) and strategic (signaling) ways to facilitate, or hinder, an observer's action recognition processes.

Keywords: coarticulation, joint action, action recognition, planning, distal actions, sequential action

OPEN ACCESS

Edited by:
Joanna Raczaszek-Leonardi, University of Warsaw, Poland

Reviewed by:
Cristina Becchio, University of Turin, Italy
Carol A. Fowler, Retired, USA

***Correspondence:**
Giovanni Pezzulo
giovanni.pezzulo@istc.cnr.it

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *29 October 2016*
Accepted: *07 February 2017*
Published: *23 February 2017*

Citation:
Donnarumma F, Dindo H and Pezzulo G (2017) Sensorimotor Coarticulation in the Execution and Recognition of Intentional Actions. Front. Psychol. 8:237. doi: 10.3389/fpsyg.2017.00237

1. INTRODUCTION

Imagine a football player who is approaching the opponent team's area with the ball. One can define the player's current goal as approaching the area, while his distal intention is passing the ball or shooting. For both his teammates and his opponents, inferring the player's distal intention (not only his current goal) offers an advance opportunity to help or hinder him, highlighting the importance of goal and intention recognition in realistic social interactions, cooperative or competitive. From a computational perspective, another's proximal goals and distal intentions can be considered *hidden* (i.e., non-observable) cognitive variables that can be inferred based on observables (e.g., the player's behavior) and prior knowledge (e.g., tactics used by the soccer team) (Wolpert et al., 2003). While generally difficult in real-world social settings, goal and intention recognition may be less formidable than commonly believed, because proximal kinematics turn out to be very informative.

A series of studies have shown that humans are surprisingly good at inferring another person's proximal goals or distal intentions, even with apparently little data (Sartori et al., 2009). One

recent study reveals that participants who observed grasping movements were able to report accurately whether the to-be-grasped object was small or big as early as 80 ms after movement onset, suggesting that action kinematics can be very informative at early perceptual stages (Ansuini et al., 2016). A similar case may be made for the recognition of distal intentions. For example, considering the case in which an agent makes a decision between "grasping a bottle to pour water" vs. "to move the bottle," evidence shows that the agent's decision is already discriminable by the first part of the motor action, i.e., the grasping of the bottle (Sartori et al., 2011). In fact, the way in which the bottle is grasped turns out to be slightly different (e.g., at the level of action kinematics) in the two cases. More in general, many studies show a "tendency to grasp objects differently depending on what one plans to do with the objects" (Rosenbaum et al., 2012), which means that hand preshape can be used as a cue to infer the distal intention. This situation has equivalents in other domains, such as for example linguistics, in which it is widely known that the pronunciation of segments depends on other segments which are close to them, e.g., the next segment (*coarticulation*, see e.g., Fowler, 1980; Fowler and Saltzman, 1993; Mahr et al., 2015). These subtle changes in the action kinematics provide information about the performer's goals (Sartori et al., 2009; Neal and Kilner, 2010; Manera et al., 2011; Becchio et al., 2012; Naish et al., 2013; Quesque et al., 2013; Ansuini et al., 2015; Lewkowicz et al., 2015; Cavallo et al., 2016). At least in some cases, even subtle cues are detectable and can help observers to infer the performer's distal intentions early on, thus resulting in communicative and not only pragmatic effects.

The informativeness of early kinematic cues may even increase in explicitly cooperative social settings. For example, one study reveals that during the same motor action of placing an object, the deceleration phase was found to be slower when a "giving" action (proximal goal) was directed to another individual than when it was performed without this social constrain (Becchio et al., 2008). A series of other studies have shown that, when engaged in social interactions, co-actors usually *signal* their intentions and carve their action kinematics in ways that make their action goals easier to discriminate, when there is asymmetric information and the performer agent is more knowledgeable than the observer about the task at hand (Vesper et al., 2010; Pezzulo, 2011; Pezzulo and Dindo, 2013; Pezzulo et al., 2013a; Sacheli et al., 2013; Candidi et al., 2015).

These and other studies have assessed the usefulness of (early) kinematic cues for understanding an actor's proximal goals but also his distal intentions. One possible explanation for this phenomenon is that, in the context of grasping actions, an object can be handled and manipulated differently depending on a performer's intention—hence the agent's intention can be inferred from the way the agent performs the motor action. This explanation, however, lacks a quantitative (or computational) characterization so far and it is unclear whether one can derive the benefits of distal intention recognition from normative principles, e.g., the minimization of action costs. Furthermore, it is unclear if the explanation is *sufficient* to explain the data; for example, if appealing to early kinematic cues can fully explain the rapidity of intention recognition found in human studies, or if it is instead necessary to appeal to additional mechanisms (e.g., sophisticated prior information or evolutionary adaptations for intention recognition that are fundamentally different from those that permit recognizing proximal action goals).

In this paper, we offer a computationally-guided explanation of distal intention recognition that is rooted in normative theories of computational motor control and (embodied) sequential action (Sandamirskaya and Schöner, 2010; Rosenbaum et al., 2012; Pezzulo et al., 2014, 2017; Lepora and Pezzulo, 2015; Pezzulo and Cisek, 2016). In a control theoretic perspective, proximal actions have to simultaneously fulfill the concurrent demands of proximal and distal goals (or first-order and higher-order planning). In other words, any goal-directed action is shaped according to its proximal and distal goals: *first-order planning* (associated with proximal goals) determines object handling grasp trajectory according to immediate task demands (e.g., tuning to the orientation or the grip size for an object to be grasped); *higher-order planning* (associated with distal goals) alters one's object manipulation behavior not only on the basis of immediate task demands but also on the basis of the next tasks to be performed. This would imply that in certain conditions one can impose a cost on the proximal action or execute it suboptimally in order to fulfill the requirements of a distal action, e.g., a waiter can grasp a glass with a thumb-down posture if he has to successively turn it upright (Rosenbaum et al., 1990). The necessity of simultaneously optimizing proximal and distal components of an action sequence (e.g., "reaching and grasping a bottle to pour water" vs. "to move the bottle") implies the coarticulation of consecutive motor acts, which would thus provide a normative rationale for the differences in the former part of the sequence ("reaching and grasping the bottle") depending on the latter part or distal intention[1].

Below we present a computational analysis of coarticulation during object grasping showing that (i) an agent who coarticulates proximal and distal actions produces different kinematic patterns in the first part of a sequential action ("reaching and grasping the bottle") depending on his distal goal ("pouring" or "moving the bottle"); (ii) in turn, coarticulation gives rise to kinematic features that are sufficient for observers to correctly discriminate the agent's distal intention early in time—at least in some cases. Our analysis provides computational-level support for the idea that accurate intention recognition may be due to early kinematic cues elicited during proximal actions, without necessarily requiring additional mechanisms. In turn, the elicitation of informative cues may be a byproduct of the optimization of sequential actions and does not need to have necessarily a social goal (e.g., facilitation of action recognition, like in signaling Vesper et al., 2010; Pezzulo, 2011; Sacheli et al., 2013)—although of course automatic (coarticulation) and

[1]For the sake of simplicity, here we equate coarticulation and assimilation, see also (Jerde et al., 2003). However, there is a conceptual difference between the two: coarticulation is the underlying process (i.e., the temporal overlap between sequential actions) while assimilation is the superficial result (in terms of increasing the similarity of the last part of the first movement to the first part of the last movement).

strategic (signaling) modulations of one's own action kinematics can be merged.

2. COMPUTATIONAL APPROACH

In computational motor control, it is widely assumed that action representations stem from (probabilistic) internal models (Wolpert et al., 2003; Jeannerod, 2006; Shadmehr and Krakauer, 2008; Friston et al., 2010, 2017; Pezzulo et al., 2015, in press; Donnarumma et al., 2016; Maisto et al., 2016; Stoianov et al., 2016). These models can be hierarchical, with higher hierarchical levels encoding more abstract and distal aspects and lower hierarchical levels encoding more proximal aspects that are related to action performance. At lower levels, actions such as grasping or pouring can be associated with probability distributions over hand kinematics (e.g., controls of angles of fingers), which of course change over time as the action unfolds.

Within this general probabilistic framework, we model the performer agent using a computational method that combines basic actions (or motor primitives) such as grasping and pouring to realize a sequential action (e.g., grasp a bottle to move it or pour), with or without coarticulating them. Furthermore, we model the observer agent using a computational method that infers the performer agent's current action, by "simulating" the execution of (the same) motor primitives for grasping, moving and pouring. Below we briefly introduce these two computational models, which we successively specify more formally.

2.1. Rationale of the Computational Approach

2.1.1. Performer Agent

According to our coarticulation hypothesis, we describe the planning of sequential actions as the *coarticulation* (or assimilation) of two successive motor primitives, e.g., motor primitive for reaching-and-grasping and one for grasping-and-pouring. Intuitively, assimilation implies that if the two sequential actions are modeled by two different probability distributions of hand kinematics (Dindo et al., 2011; Pezzulo et al., 2013a), these two distributions are made more similar by sampling from their probabilistic superposition (aka, coarticulated distribution) rather than the two original distributions, over time. **Figure 1** offers a schematic illustration of this concept in a simplified 2D domain, where the proximal action (from time zero to time 1,000) corresponds to moving a mouse to the center-right, and the distal action (from time 1,000 to time 2,000) corresponds to moving the mouse to the top-right or bottom-right. The colors correspond to the mean and variance of the probability distributions of hypothetical center-right, top-right and bottom-right mouse movements. The figure shows how the same proximal action—move to center-right—can be either independent from (top panel) or assimilated/coarticulated with (bottom panel) the successive action of reaching the top-right or bottom-right. In the latter case, the effects of assimilation on the mouse movements are apparent from time 600, well before the (theoretical) beginning of the distal action.

2.1.2. Observer Agent

According to motor theories of cognition, the computational mechanisms (and internal models) used for action planning and execution are also reused for action understanding, using *motor simulation* (Jeannerod, 2006). In keeping with this idea, we model the action observation process as a (probabilistic) inference problem, in which an observer agent considers multiple possible hypotheses that correspond to the actions that may have generated the observed sensory stimuli (i.e., whether the performer agent is grasping for pouring vs. grasping for moving) and has to select one of them. To do so, the observer agent simulates executing multiple actions in parallel (from his own motor repertoire), compares the predictions under these different hypotheses with the observed movements, and assigns high probability to the action/hypothesis that generates the smaller prediction error. This process is iterated over time using a probabilistic scheme (see below), so that as the performer agent's actions unfold in time, evidence accumulates for one of the alternative hypotheses. Note that using this framework implicitly requires the assumption that performer and observer agents share the same set of motor primitives, although the probabilistic parameters might differ according to individual actor's knowledge and expertise. Our simulations will show that this motor simulation process converges more readily to the correct explanation when the performer agent uses coarticulation—and in this latter case, an observer agent can correctly recognize the distal intention of a demonstrating agent while he is still executing the proximal action.

2.2. Formal Aspects of the Computational Approach

2.2.1. Performer Agent and the Coarticulation Distribution

Coarticulation is the process of altering one's own behavior to facilitate the next action. In this framework, a proximal action is coarticulated (or assimilated) with the next action in a sequence if the differences between the (probability distributions denoting the) two actions are minimized, while at the same time it maintains its correct pragmatics (e.g., a reaching action has to effectively reach the bottle despite being coarticulated with a successive grasping action).

To exemplify this concept, let's consider two actions (e.g., reaching a bottle and executing a power grasp), each implemented as a motor primitive (m_1 or m_2) that, for every moment in time, can be associated to a probability distribution (for example, a Gaussian distribution over its corresponding kinematic parameters such as hand and finger configurations). **Figure 2** shows the distributions associated to the two motor primitives, $p(x_t|m_1)$ for model m_1 (e.g., reaching the bottle, in blue) and $p(x_t|m_2)$ for model $m2$ (e.g., power grasp, in red), at time t. Based on these two *original* distributions, it is possible to compute the novel *coarticulated distribution* $p^{coa}(x_t|m_1)$ (e.g., reaching the bottle while preparing to grasp it with a power grasp, in orange), which corresponds to the fact that at time t, the motor primitive m_1 is *coarticulated* with m_2. Obviously, this example only describes what happens in a single temporal instant, while

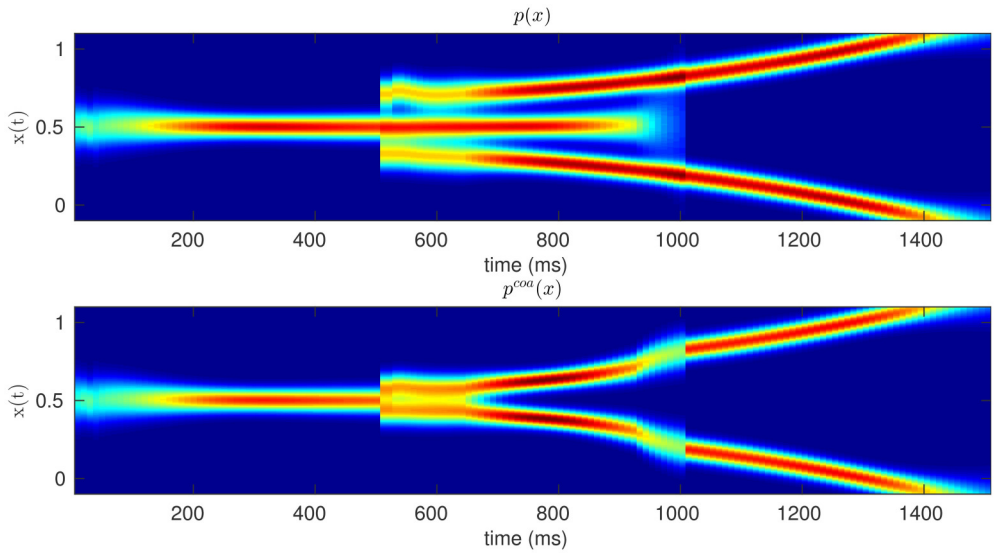


FIGURE 1 | Original (Top) vs. Coarticulated **(Bottom)** distributions for a simplified action sequence (e.g., mouse movement). The sequence consists of two parts: moving to the center-right (from zero to 1,000 ms) and moving to the top-right or bottom-right (from 1,000 to 1,500 ms); see main text for explanation. The colors denote the probability of occupying a given position in space, during time, from red (highest probability) to blue (lowest probability).

actions unfold in time. To model temporal dynamics of motor primitives, it is possible to extend the same formalism using continuous distributions, see below.

It is important to remark that any sample drawn from the coarticulation distribution (in orange) at time t should simultaneously satisfy two constraints: it should be representative of the original distribution of the first motor primitive $p(\mathbf{x}_t|m_1)$ while at the same time should have a high probability of belonging to the second motor primitive m_2 (or in more general cases, even to a set of future motor primitives, m_j). In keeping, to obtain an approximation of the coarticulation distribution, we adopt a *rejection sampling* technique. Let $\hat{\mathbf{x}}_t$ be a sample from the original distribution $p(\mathbf{x}_t|m_i)$ or a motor primitive m_i. Given K random values, $u_k \in [0, 1]$, sampled from the uniform distribution over $[0, p(\mathbf{x}_t|m_k)/p_k^{max}]$, we decide to *accept* the sample $\hat{\mathbf{x}}_t$ if the following holds:

$$u_i < w_i \cdot p(\hat{\mathbf{x}}_t|m_i) \quad \texttt{and} \quad u_j < w_j \cdot p(\hat{\mathbf{x}}_t|m_j), \forall j \neq i \qquad (1)$$

where $\mathbf{w} = [w_1, w_2, \ldots, w_K]$ is a vector of weights that modulate the contribution of the individual motor primitives in the coarticulation distribution. Intuitively, this implies that a sample is accepted if and only if it is a "good exemplar" of both (say) the grasping and the pouring distributions—implying that the novel coarticulation distribution combines aspects of grasping and pouring.

In the case of continuous distributions $p(\mathbf{x}_t|m_j)$, the *coarticulation* distribution becomes:

$$p^{coa}(\mathbf{x}_t|m_i; \mathbf{w}) \propto w_i \cdot p(\mathbf{x}_t|m_i) \prod_{j \neq i} (w_j \cdot p(\mathbf{x}_t|m_j)) \qquad (2)$$

The resulting coarticulation distribution $p^{coa}(x_t|m_i)$ is constructed in such a way that its kinematic parameters are the most probable for the motor primitive m_i but also the most similar to those of the primitive(s) to be executed next (m_j). As illustrated in **Figure 1**, the motor primitives for (say) grasping and pouring then mesh coherently over time (bottom panel: coarticulation), rather than being simply executed one after the other (top panel: no coarticulation). Another way to appreciate the key features of the coarticulation distribution is contrasting it with its "converse": the *signaling* distribution, see **Figure 2**. While the coarticulation distribution is constructed in such a way to emphasize the similarities between two motor primitives, the signaling distribution is constructed in such a way to emphasize their differences—hence the former (coarticulation) distribution is more appropriate to model assimilation effects (e.g., between two consecutive motor primitives as in our examples) and the latter (signaling) distribution is more appropriate to model dissimilation effects such as those arising during non-verbal, sensorimotor communication (Vesper et al., 2010; Pezzulo, 2011; Pezzulo and Dindo, 2013; Pezzulo et al., 2013a,b; Sacheli et al., 2013; Candidi et al., 2015). See the Appendix for a more formal treatment of the signaling distribution.

2.2.2. Observer Agent and Probabilistic Motor Simulation

Our implementation of action understanding via motor simulation is based on a Dynamic Bayesian Network (DBN) shown in **Figure 3**. DBNs are Bayesian networks representing temporal probability models in which directed arrows depict assumptions of conditional (in)dependence between variables (circles) (Murphy, 2002). Shaded nodes represent observed variables while the others are hidden and need to be

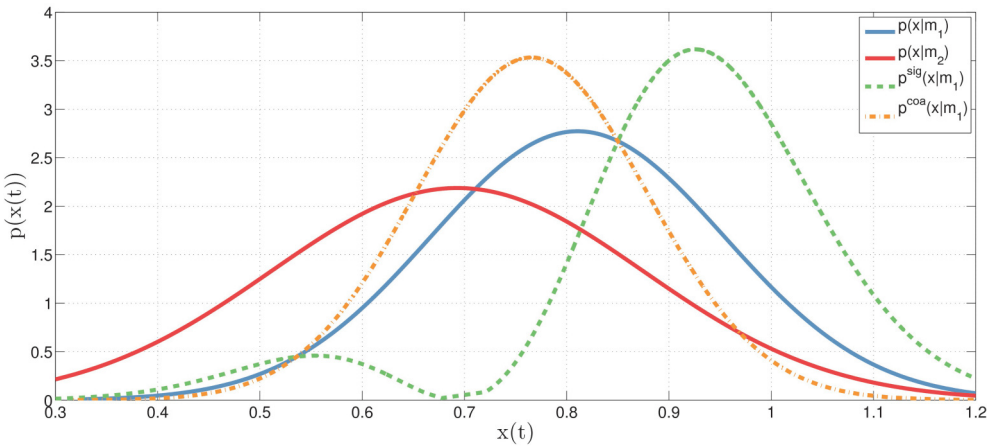


FIGURE 2 | Illustration of original, coarticulation and signaling distributions. The original (Gaussian) distributions (at time *t*) corresponding to two motor primitives: $p(x_t|m_1)$ for motor primitive m_1 (blue) and $p(x_t|m_2)$ for motor primitive *m*2 (red). In the *coarticulation distribution* $p^{coa}(x_t|m_1)$ (orange), the motor primitive m_1 is coarticulated (or assimilated, i.e., made as similar as possible) with the motor primitive m_2 using Equation (2)—where m_1 and m_2 may correspond, for example, to the first and second action in a sequence. In the *signaling distribution* $p^{sig}(x_t|m_1)$ (green), the motor primitive m_1 is dissimilated (i.e., made as different as possible) from the motor primitive m_2. See the main text for explanation.

estimated through the process of probabilistic inference. In our representation, the process of action understanding is influenced by the following factors expressed as stochastic variables in the model (see Dindo et al., 2011 for a more detailed account of the model):

1. *c*: discrete context variable;
2. *i*: index of the agent's own repertoire of goal-directed actions (proximal or distal): each action directly influences the activation of related forward and *inverse* models;
3. *u*: control-related continuous variable (e.g., forces, velocities, . . .);
4. *x*: state (e.g., the position of the demonstrator's end-effector in an allocentric reference frame);
5. *z*: observation, a perceptual measurement related to the state (e.g., the perceived position of the demonstrator's end-effector on the retina).

During action observation, the model has to infer which action the performer agent is performing (e.g., whether he or she is currently grasping, pouring, lifting, etc.—where each action, proximal or distal, is denoted by an index i_t). The goal-directed action is considered to be hidden (i.e., not directly observable); but it can be inferred on the basis of noisy sensory observations (denoted as z_t), e.g., the performer's hand movements. The logic is the usual of (inverse) Bayesian inference, which considers multiple potential actions as candidate explanations, which compete to explain the sensory data (Wolpert et al., 2003; Demiris and Khadhouri, 2005; Dindo et al., 2011; Friston et al., 2011; Pezzulo, 2013; Donnarumma et al., in press). Each action i_t is associated with a paired inverse-forward model (see Equation 4 below). Re-enacting these actions "in simulation" generates a motor control u_t (given the hidden state x_{t-1}, aka inverse model), and a prediction of the next hidden state x_t (given the motor control u_t and the previous state x_{t-1}, aka forward model). Comparing the predicted and sensed movements under various competing hypotheses (e.g., grasping, pouring) permits to assess which one generates less prediction error, hence explaining better the data. A priori contextual information c_t can bias the inferential process and the initial choice of the internal models to test (in case they are too numerous).

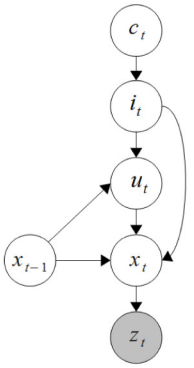
The following equations describe the observation model (Equation 3), which specifies the way (noisy) sensory stimuli are used to estimate the state of the demonstrator (e.g., hand position); the transition model (Equation 4), which specifies how the state of the demonstrator evolves as a function of his or her goals and motor commands; and the a priori distribution over the set of hidden variables (Equation 5), which represents the perceiver's prior belief and is a necessary ingredient of Bayesian systems.

$$p(\mathcal{Z}_t|\mathcal{X}_t) = p(z_t|x_t) \tag{3}$$

$$p(\mathcal{X}_t|\mathcal{X}_{t-1}) = p(x_t|x_{t-1}, u_t, i) \cdot p(u_t|x_{t-1}, i) \tag{4}$$

$$p(\mathcal{X}_0) = p(x_0) \cdot p(c_0) \cdot p(i|c_0) \tag{5}$$

The inference exploits the usual (prediction) error-correction mechanisms of Bayesian systems. The model starts with prior hypotheses about the demonstrator's actions and intentions, and these are iteratively revised as new sensory evidence is sampled. The evidence provided by the perceptual process and the observed states (z_t) is responsible for "correcting" the posterior distribution when integrating the observation model $p(z_t|x_t)$. In other words, those parts of the hidden state that are in accordance with the observations will exhibit peaks in the posterior distribution. Since those states have been produced by a goal-directed motor primitive, the marginalization of the final



Graphical model

c_t	context	$[0, ..., C]$
i	goal-directed action index	$[0, ..., I]$
u_t	motor act or control	*continuous*
x_t	state	*continuous*
z_t	observed state	*continuous*

Stochastic variables

$p(c_t)$	contextual information
$p(i_t\|c_t)$	contextual induction
$p(u_t\|x_{t-1}, i)$	inverse model
$p(x_t\|x_{t-1}, u_t, i)$	forward model
$p(z_t\|x_t)$	observation model (prediction error)

Probability densities

FIGURE 3 | Graphical model (DBN) for action simulation based on coupled forward-inverse models. See main text and Dindo et al. (2011) for details.

posterior distribution produces the required discrete distribution over motor primitives, $p(i_t|z_{1:t})$.

Thus, the motor primitive with the highest probability (above a fixed threshold) is selected as the "winning" primitive; such an inference process can be iterated over time by using the full posterior distribution as the prior for the next step, until convergence. Ultimately, the aim of the whole process is estimating the probability of each model given the current observations so far (i.e., likelihood). The most plausible model is the one that maximizes the posterior probability of the model. As usual in a Bayesian setting, the whole process is influenced by the choice of the prior distributions for the available motor primitives: the more likely is a particular motor primitive a-priori, the more reliable and fast its recognition. In particular, using this framework action recognition is influenced by an auxiliary (contextual) variable, which can intuitively reflect an agent's contextual knowledge (e.g., that pouring is highly unlikely if the bottle is almost empty) that biases the motor primitives that are actually simulated by the agent. While prior probabilities and contextual information are important in real-life scenarios, we do not use them in our simulations.

3. EXPERIMENTAL SETUP AND RESULTS

We performed a series of computational simulations, in which one (performer) agent executes one of two sequential actions (e.g., "reaching and grasping a bottle to pour water" vs. "reaching and grasping a bottle to move it") in two conditions: with and without the coarticulation method explained in Section 2.2.1. At the same time, the other (observer) agent has to disambiguate these two alternatives as soon and as accurately as possible, using the probabilistic motor simulation methods introduced in Section 2.2.2. These simulations permit us to study the benefits of coarticulation, and to test the "sufficiency" hypothesis introduced earlier: namely, that kinematic features at early stages of a coarticulated action permit an observer to recognize the action. In our scenario, this means that a sequential action (e.g., "reaching and grasping a bottle to pour water") can be discriminated already during the first (reaching) phase. Conversely, when the same action is executed without coarticulation, it can only be recognized during the second phase, after the agent has grasped the bottle.

In our simulations, for the sake of simplicity we focused on two two-step sequential actions: reach-and-pour vs. reach-and-move. In practice, we used three motor primitives: the former primitive (reach-to-grasp) corresponds to the first step in both sequences, while the other two primitives (grasp-to-pour and grasp-to-move) correspond to the two final actions to complete the first and second sequential actions, respectively. At each moment in time, from 0 ms (beginning of sequential action) to 1,500 ms (end of sequential action), each motor primitive corresponds to a probability distribution over controls of finger, thumb and wrist of a (human) hand.

The motor primitives were derived based on controls and parameters extracted from human data collected from six adult male participants. Each participant executed every primitive action 50 times, and data on angles of finger, thumb and wrist were collected using a dataglove (HumanGlove - Humanware S.r.l., Pontedera, Pisa, Italy) endowed with 16 sensors. The former (reach-to-grasp) motor primitive was acquired while participants reached an object the size of a bottle with a concave constriction (see also Sartori et al., 2011), with no knowledge of the next action to perform with it. We selected the latter two primitives (grasp-to-pour and grasp-to-move) as instances of power grasp and a precision grip actions, respectively, in which the end-position of the fingers was analogous to the positions reported by Sartori et al. (2011) while humans grasped a bottle to pour or move it, respectively.

The internal dynamical models (motor primitives) used in the simulations were obtained by regressing the aforementioned data (50 trials for 6 participants for each primitive), to obtain probability distributions over angles of finger, thumb and wrist, over time. For each motor primitive, we used an *Echo State Gaussian Process* (ESGP) (Chatzis and Demiris, 2011), a method for the Bayesian modeling of sequential data that produces a measure of confidence (or uncertainty) on the generated predictions (the model predictive density), which can be directly used within our computational approach.

In the simulations reported below, a non-coarticulated action corresponds to the first primitive (reach-to-grasp) being used for the first 1,000 ms, while one of the two remaining primitives (grasp-to-pour or grasp-to-move, depending on the task) is used for the successive 500 ms. A coarticulated action corresponds to the first primitive (reach-to-grasp) being coarticulated with one of the two remaining primitives (grasp-to-pour or grasp-to-move, depending on the task) during the interval between 500 and 1,000 ms, using the coarticulation method explained in Section 2.2.1. In other words, we derive the coarticulated actions by "meshing" two primitives, not by using separate ESGPs. Note that in the simulations, we coarticulated the index finger and thumb controls (not the wrist controls), coherent with their importance in grasping and pouring actions (Sartori et al., 2011).

A first result of our simulations is that during the execution of the former (reach-to-grasp) motor primitive in the sequence, Maximum Grip Aperture and Time of Maximum Grip Aperture differ significantly if the primitive is coarticulated with a grasp-to-pour primitive, with a grasp-to-move primitive, or not coarticulated at all, see **Figure 4**. This result is not remarkable *per se*, but can be considered as a "safety check" that the different intention elicits different action kinematics, with a pattern that is qualitatively coherent with the results of Sartori et al. (2011) in a similar scenario. What is more important for us was studying whether (and how) this difference in action kinematics translates into an advantage for the observer agent, at early stages of the performer's agent movement.

To answer this question, we simulated the behavior of an observer agent that has to recognize the actions performed by the performer agent, using the probabilistic motor simulation mechanism described in Section 2.2.2. As shown in **Figure 5**, the observer agent had a clear advantage in recognizing the performed action when it was coarticulated. More specifically, the figure shows that without coarticulation the performer agent's distal intention (pouring or moving the bottle) can be recognized only after he reaches the bottle (i.e., after time 1,000), while with coarticulation it can be recognized much earlier, during the execution of the first motor primitive (i.e., well before time 1,000). This latter result illustrates that coarticulation affords intention recognition in ways that are qualitatively different from the mere execution of a (non-informed) action.

4. DISCUSSION

Humans excel at recognizing distal intentions on the basis of (apparently) little information, but the cognitive and computational mechanisms underlying this ability are incompletely known. We have proposed that normative principles regulating the coarticulation of sequential actions can explain how it is possible to infer a performer's distal intention by looking at the kinematics of his proximal actions.

To test this idea, we implemented a series of simulations in which the performer agent executes sequential actions (reach-and-pour or reach-and-move) as sequences of two primitives (reach-to-grasp and grasp-to-pour, or reach-to-grasp and grasp-to-move) with or without coarticulation. Our results show that two successive actions can be coarticulated (or assimilated) in such a way that the kinematics of the proximal action are adequate for (and informative of) the next action(s) in the sequence. Indeed, our results show that, first, coarticulated actions have characteristic kinematic features compared to non-coarticulated actions, and second, that these features may be *sufficient* for an observer agent to correctly recognize the performer's agent distal intention early on. This result holds despite the fact that we used simplified motor primitives and only coarticulated index finger and thumb controls. In principle, an observer agent having access to richer visual stimuli and more sophisticated primitives (with more controls and degrees of freedom) may enjoy additional benefits; it is however possible that coarticulation only operates on a restricted set of degrees of freedom, e.g., those that are necessary to solve the task, as for the uncontrolled manifold hypothesis (Scholz and Schöner, 1999). At the same time, it is possible that in real-life conditions, some information encoded in movement kinematics that would be potentially useful to infer a performer's intention may nevertheless remain invisible to observers—for example, when parametric variations are too small to be detected (Naish et al., 2013; Cavallo et al., 2016). Our computational study shows that coarticulation promotes the appropriate preconditions for advance intention understanding, but the additional factors that may favor (or prevent) an advantage for observer agents remain to be fully assessed.

Our emphasis on the sufficiency of kinematic cues to solve intention recognition tasks does not imply that interactive agents do not use other sources of information, such as (prior information on) the context in which the action takes place. For example, it has been argued that the same action (approaching a person with a knife) can be motivated by two different intentions (e.g., Dr. Jekyll who wants to cure or Mr. Hyde who wants to kill) and these can be disambiguated based on the place where the action occurs (operating room or dark street) (Kilner et al., 2007), but see Jacob and Jeannerod (2005); Kilner et al. (2007); Becchio et al. (2008) for alternative proposals. This kind of prior information can be readily incorporated in the action recognition scheme described above, through the contextual (*C*) node of the DBN. By considering the probabilistic relations between contexts and actions, it would be possible to bias the action recognition process and distinguish the intentions motivating two actions, even when they are kinematically identical—a situation that, as we have discussed, may be more the exception than the rule. Furthermore, it would be possible to extend the model discussed here so that it also directs saccadic eye movements to the most informative parts of the demonstrator's actions, in keeping with the idea that action recognition uses an active sensing scheme (Donnarumma et al., in press). Modeling eye movements would help understanding under which conditions subtle kinematic cues that are embedded in goal-directed actions are detected by observer agents.

Following a motor cognition approach, our model implements action recognition as a (Bayesian) inferential process that uses the logic of "analysis-by-synthesis" or action simulation (Jeannerod, 2006). This is in keeping with evidence (reviewed in Grafton, 2009) that observer agents simulate the

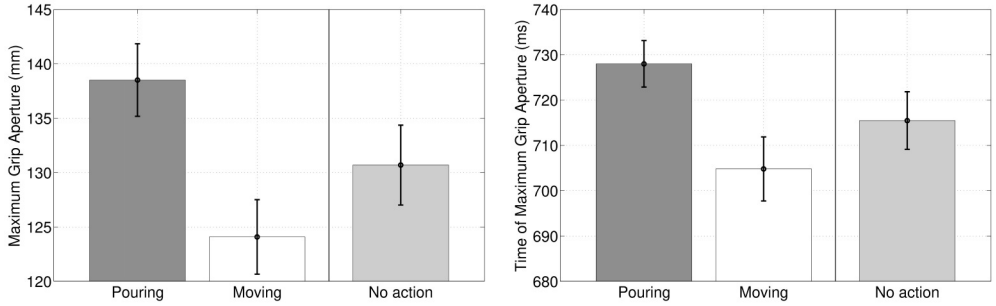


FIGURE 4 | Maximum Grip Aperture and Time of Maximum Grip Aperture when (1) the reach-to-grasp primitive is coarticulated with grasp-to-pour, (2) the reach-to-grasp primitive is coarticulated with grasp-to-move and (3) the reach-to-grasp primitive is not coarticulated (baseline condition), as if there was no successive action.

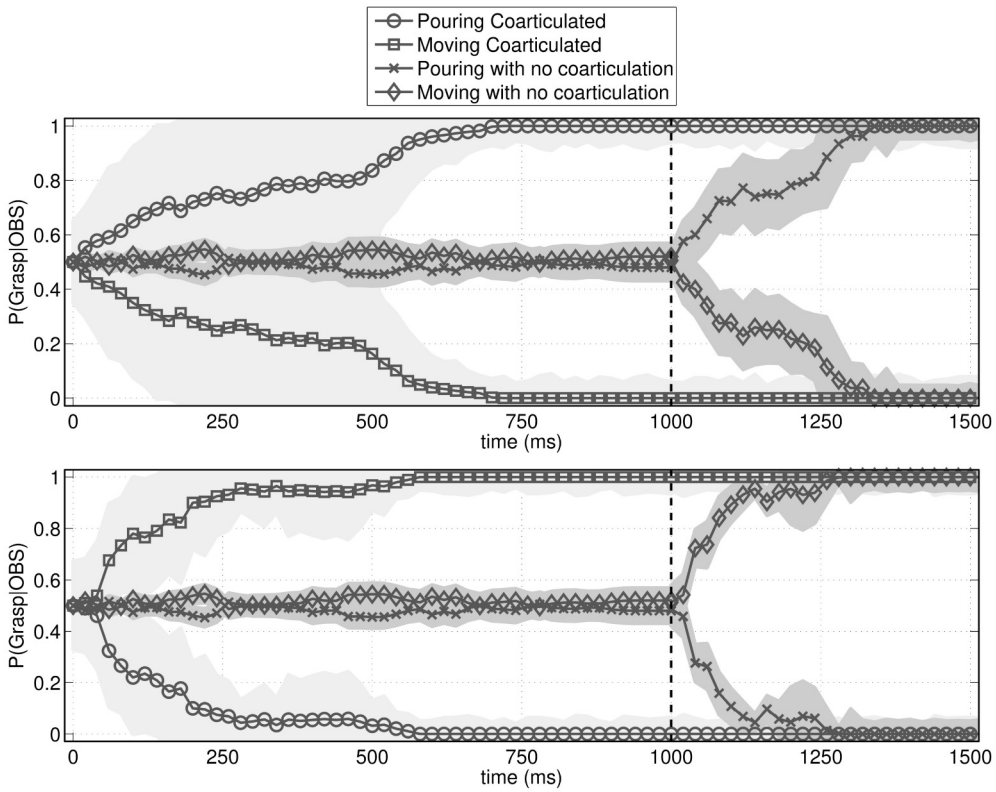


FIGURE 5 | Probability assigned by the observer agent to the to-be-recognized action (pouring or moving). The figures show the mean probability and standard deviation (shaded area) of pouring vs. moving action sequences given the current observations. The vertical dotted bar is the moment when the performer agent reaches the bottle. **(Top)** Probabilities over time when the to-be-recognized action is pouring. **(Bottom)** Probabilities over time when the to-be-recognized action is moving. In both cases, if the action is not coarticulated, it is recognized only after the bottle is reached. Instead, if the action is coarticulated, it is recognized early on, during the execution of the proximal action.

actions they observe in their brains. Alternative hypotheses point, for example, to a more ecological or enactive view of action understanding, which appeal to "direct perception" rather than (Bayesian) inference (Gibson, 1966). While this alternative perspective would differ from our implementation, the logic of our argument here may be the same—that is, that coarticulation generates information that an observer agent can use to form an advance understanding of the performer's goals (via Bayesian inference or direct perception).

It is notable that we have illustrated the model by discussing coarticulation in the domain of reaching and grasping actions, where essentially coarticulation implies the preshaping of hands before executing a grasping action (Jeannerod, 2006). However, the phenomenon of coarticulation is evident in all sequential

actions, and the model presented here is (in principle) general enough to address analogous phenomena in other domains, including speech, sign language (Jerde et al., 2003) and the planning of smooth action sequences (Rosenbaum et al., 2006). It remains to be assessed by future studies whether the computational scheme presented here is empirically adequate to explain sequential action in these and other domains, or if it needs to be extended to include more sophisticated internal generative models (e.g., of hierarchical dynamics rather than only sequences of motor primitives Kiebel et al., 2008, 2009; Donnarumma et al., 2015a,b)—as well as the relative merits of alternative frameworks such as those stemming from a dynamical systems perspective (Kelso, 1995; Marsh et al., 2006, 2009).

To sum up, according to this (normative) proposal, the main goal of coarticulation is to optimize sequential actions, and the facilitatory effects for social cognition are byproducts of this process. In other words, according to this proposal, there is no need of any action recognition or mindreading adaptation in the observer, because the action recognition process is greatly facilitated by the performer—albeit often unwittingly (but see the Appendix). This process is effective because during the execution of sequential actions, there is a sort of backward influence from the latter action (and its constraints) to the former action. Thus, the former action already includes subtle but reliable kinematic cues, which can be used to infer the performer's distal goal—and we humans excel at picking up these cues.

AUTHOR CONTRIBUTIONS

FD, HD, and GP conceived the study, collected and analyzed data, and wrote the manuscript.

FUNDING

The present research is funded by the Human Frontier Science Program (HFSP), award number RGY0088/2014, by the EU's FP7 under grant agreement no FP7-ICT-270108 (Goal-Leaders).

ACKNOWLEDGMENTS

The GEFORCE Titan used for this research was donated by the NVIDIA Corporation.

REFERENCES

Ansuini, C., Cavallo, A., Bertone, C., and Becchio, C. (2015). Intentions in the brain the unveiling of mister hyde. *Neuroscientist* 21, 126–135. doi: 10.1177/1073858414533827

Ansuini, C., Cavallo, A., Koul, A., D'Ausilio, A., Taverna, L., and Becchio, C. (2016). Grasping others' movements: Rapid discrimination of object size from observed hand movements. *J. Exp. Psychol. Hum. Percept. Perform.* 42, 918–929. doi: 10.1037/xhp0000169

Becchio, C., Manera, V., Sartori, L., Cavallo, A., and Castiello, U. (2012). Grasping intentions: from thought experiments to empirical evidence. *Front. Hum. Neurosci.* 6:117. doi: 10.3389/fnhum.2012.00117

Becchio, C., Sartori, L., Bulgheroni, M., and Castiello, U. (2008). The case of Dr. Jekyll and Mr. Hyde: a kinematic study on social intention. *Conscious. Cogn.* 17, 557–564. doi: 10.1016/j.concog.2007.03.003

Candidi, M., Curioni, A., Donnarumma, F., Sacheli, L. M., and Pezzulo, G. (2015). Interactional leader–follower sensorimotor communication strategies during repetitive joint actions. *J. R. Soc. Inter.* 12:20150644. doi: 10.1098/rsif.2015.0644

Cavallo, A., Koul, A., Ansuini, C., Capozzi, F., and Becchio, C. (2016). Decoding intentions from movement kinematics. *Sci. Rep.* 6:37036. doi: 10.1038/srep37036

Chatzis, S. P., and Demiris, Y. (2011). Echo state gaussian process. *IEEE Trans. Neural Netw.* 22, 1435–1445. doi: 10.1109/TNN.2011.2162109

Demiris, Y., and Khadhouri, B. (2005). Hierarchical attentive multiple models for execution and recognition (hammer). *Robot. Autonom. Syst. J.* 54, 361–369. doi: 10.1016/j.robot.2006.02.003

Dindo, H., Zambuto, D., and Pezzulo, G. (2011). "Motor simulation via coupled internal models using sequential monte carlo," in *Proceedings of IJCAI 2011* (Barcelona), 2113–2119.

Donnarumma, F., Prevete, R., De Giorgio, A., Montone, G., and Pezzulo, G. (2015a). Learning programs is better than learning dynamics: a programmable neural network hierarchical architecture in a multi-task scenario. *Adapt. Behav.* 24, 27–51. doi: 10.1177/1059712315609412

Donnarumma, F., Costantini, M., Ambrosini, E., Friston, K., and Pezzulo, G. (in press). Action perception as hypothesis testing. *Cortex*. doi: 10.1016/j.cortex.2017.01.016

Donnarumma, F., Maisto, D., and Pezzulo, G. (2016). Problem solving as probabilistic inference with subgoaling: explaining human successes and pitfalls in the tower of hanoi. *PLoS Comput. Biol.* 12:e1004864. doi: 10.1371/journal.pcbi.1004864

Donnarumma, F., Prevete, R., Chersi, F., and Pezzulo, G. (2015b). A programmer-interpreter neural network architecture for prefrontal cognitive control. *Int. J. Neural. Syst.* 25:1550017. doi: 10.1142/S0129065715500173

Fowler, C. A. (1980). Coarticulation and theories of extrinsic timing. *J. Phonet.* 8, 113–133.

Fowler, C. A., and Saltzman, E. (1993). Coordination and coarticulation in speech production. *Lang. Speech* 36, 171–195.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., and Pezzulo, G. (2017). Active inference: a process theory. *Neural Comput.* 21, 1–49. doi: 10.1162/NECO_a_00912

Friston, K., Mattout, J., and Kilner, J. (2011). Action understanding and active inference. *Biol. Cybern.* 104, 137–160. doi: 10.1007/s00422-011-0424-z

Friston, K. J., Daunizeau, J., Kilner, J., and Kiebel, S. J. (2010). Action and behavior: a free-energy formulation. *Biol. Cybern.* 102, 227–260. doi: 10.1007/s00422-010-0364-z

Gibson, J. (1966). *The Senses Considered as Perceptual Systems*. Boston, MA: Houghton Mifflin.

Gonzalez, D. A., Studenka, B. E., Glazebrook, C. M., and Lyons, J. L. (2011). Extending end-state comfort effect: do we consider the beginning state comfort of another? *Acta Psychol. (Amst.)* 136, 347–353. doi: 10.1016/j.actpsy.2010.12.009

Grafton, S. T. (2009). Embodied cognition and the simulation of action to understand others. *Ann. N.Y. Acad. Sci.* 1156, 97–117. doi: 10.1111/j.1749-6632.2009.04425.x

Jacob, P., and Jeannerod, M. (2005). The motor theory of social cognition: a critique. *Trends Cogn. Sci.* 9, 21–25. doi: 10.1016/j.tics.2004.11.003

Jeannerod, M. (2006). *Motor Cognition*. New York, NY: Oxford University Press.

Jerde, T. E., Soechting, J. F., and Flanders, M. (2003). Coarticulation in fluent fingerspelling. *J. Neurosci.* 23, 2383–2393. Available online at: http://www.jneurosci.org/content/23/6/2383

Kelso, J. (1995). *Dynamic Patterns*. Cambridge, MA; Bradford: MIT Press.

Kiebel, S. J., Daunizeau, J., and Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Comput. Biol.* 4:e1000209. doi: 10.1371/journal.pcbi.1000209

Kiebel, S. J., Daunizeau, J., and Friston, K. J. (2009). Perception and hierarchical dynamics. *Front. Neuroinform.* 3:20 doi: 10.3389/neuro.11.020.2009

Kilner, J. M., Friston, K. J., and Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. *Cogn. Process.* 8, 159–166. doi: 10.1007/s10339-007-0170-2

Lepora, N. F., and Pezzulo, G. (2015). Embodied choice: how action influences perceptual decision making. *PLoS Comput. Biol.* 11:e1004110. doi: 10.1371/journal.pcbi.1004110

Lewkowicz, D., Quesque, F., Coello, Y., and Delevoye-Turrell, Y. N. (2015). Individual differences in reading social intentions from motor deviants. *Front. Psychol.* 6:1175. doi: 10.3389/fpsyg.2015.01175

Mahr, T., McMillan, B. T., Saffran, J. R., Weismer, S. E., and Edwards, J. (2015). Anticipatory coarticulation facilitates word recognition in toddlers. *Cognition* 142, 345–350. doi: 10.1016/j.cognition.2015.05.009

Maisto, D., Donnarumma, F., and Pezzulo, G. (2016). Nonparametric problem-space clustering: learning efficient codes for cognitive control tasks. *Entropy* 18, 61. doi: 10.3390/e18020061

Manera, V., Becchio, C., Cavallo, A., Sartori, L., and Castiello, U. (2011). Cooperation or competition? discriminating between social intentions by observing prehensile movements. *Exp. Brain Res.* 211, 547–556. doi: 10.1007/s00221-011-2649-4

Marsh, K. L., Richardson, M. J., Baron, R. M., and Schmidt, R. (2006). Contrasting approaches to perceiving and acting with others. *Ecol. Psychol.* 18, 1–38. doi: 10.1207/s15326969eco1801_1

Marsh, K. L., Richardson, M. J., and Schmidt, R. C. (2009). Social connection through joint action and interpersonal coordination. *Topics Cogn. Sci.* 1, 320–339. doi: 10.1111/j.1756-8765.2009.01022.x

Murphy, K. P. (2002). *Dynamic Bayesian Networks: Representation, Inference and Learning*. Ph.D. thesis, UC Berkeley, Computer Science Division.

Naish, K. R., Reader, A. T., Houston-Price, C., Bremner, A. J., and Holmes, N. P. (2013). To eat or not to eat? kinematics and muscle activity of reach-to-grasp movements are influenced by the action goal, but observers do not detect these differences. *Exp. Brain Res.* 225, 261–275. doi: 10.1007/s00221-012-3367-2

Neal, A., and Kilner, J. M. (2010). What is simulated in the action observation network when we observe actions? *Eur. J. Neurosci.* 32, 1765–1770. doi: 10.1111/j.1460-9568.2010.07435.x

Pezzulo, G. (2011). Shared representations as coordination tools for interactions. *Rev. Philos. Psychol.* 2, 303–333. doi: 10.1007/s13164-011-0060-5

Pezzulo, G. (2013). Studying mirror mechanisms within generative and predictive architectures for joint action. *Cortex* 49, 2968–2969. doi: 10.1016/j.cortex.2013.06.008

Pezzulo, G., and Cisek, P. (2016). Navigating the affordance landscape: feedback control as a process model of behavior and cognition. *Trends Cogn. Sci.* 20, 414–424. doi: 10.1016/j.tics.2016.03.013

Pezzulo, G., and Dindo, H. (2013). Intentional strategies that make co-actors more predictable: the case of signaling. *Behav. Brain Sci.* 36, 43–44. doi: 10.1017/S0140525X12002816

Pezzulo, G., Donnarumma, F., and Dindo, H. (2013a). Human sensorimotor communication: a theory of signaling in online social interactions. *PLoS ONE* 8:e79876. doi: 10.1371/journal.pone.0079876

Pezzulo, G., Iodice, P., Ferraina, S., and Kessler, K. (2013b). Shared Action Spaces: a basis function framework for social re-calibration of sensorimotor representations supporting joint action *Front. Hum. Neurosci.* 7:800. doi: 10.3389/fnhum.2013.00800

Pezzulo, G., Iodice, P., Donnarumma, F., Dindo, H., and Knoblich, G. (in press). Avoiding accidents at the champagne reception: a study of joint lifting and balancing. *Psychol. Sci.* doi: 10.1177/0956797616683015

Pezzulo, G., Kemere, C., and van der Meer, M. (2017). Internally generated hippocampal sequences as a vantage point to probe future-oriented cognition. *Ann. N.Y. Acad. Sci.* [Epub ahead of print]. doi: 10.1111/nyas.13329

Pezzulo, G., Rigoli, F., and Friston, K. (2015). Active inference, homeostatic regulation and adaptive behavioural control. *Prog. Neurobiol.* 134, 17–35. doi: 10.1016/j.pneurobio.2015.09.001

Pezzulo, G., van der Meer, M. A., Lansink, C. S., and Pennartz, C. M. A. (2014). Internally generated sequences in learning and executing goal-directed behavior. *Trends Cogn. Sci.* 18, 647–657. doi: 10.1016/j.tics.2014.06.011

Quesque, F., Lewkowicz, D., Delevoye-Turrell, Y. N., and Coello, Y. (2013). Effects of social intention on movement kinematics in cooperative actions. *Front. Neurorobot.* 7:14. doi: 10.3389/fnbot.2013.00014

Rosenbaum, D., Marchak, F., Barnes, H., Vaughan, J., Slotta, J., and Jorgensen, M. (1990). "Constraints for action selection: overhand versus underhand grips," in *Attention and Performance XIII. Motor Representation and Control*, ed M. Jeannerod (Hillsdale: Lawrence Erlbaum), 211–265.

Rosenbaum, D. A., Chapman, K. M., Weigelt, M., Weiss, D. J., and van der Wel, R. (2012). Cognition, action, and object manipulation. *Psychol. Bull.* 138, 924. doi: 10.1037/a0027839

Rosenbaum, D. A., Halloran, E. S., and Cohen, R. G. (2006). Grasping movement plans. *Psychon. Bull. Rev.* 13, 918–922. doi: 10.3758/BF03194019

Sacheli, L. M., Tidoni, E., Pavone, E. F., Aglioti, S. M., and Candidi, M. (2013). Kinematics fingerprints of leader and follower role-taking during cooperative joint actions. *Exp. Brain Res.* 226, 473–486. doi: 10.1007/s00221-013-3459-7

Sandamirskaya, Y., and Schöner, G. (2010). An embodied account of serial order: how instabilities drive sequence generation. *Neural Netw.* 23, 1164–1179. doi: 10.1016/j.neunet.2010.07.012

Sartori, L., Becchio, C., Bara, B. G., and Castiello, U. (2009). Does the intention to communicate affect action kinematics? *Conscious. Cogn.* 18, 766–772. doi: 10.1016/j.concog.2009.06.004

Sartori, L., Straulino, E., and Castiello, U. (2011). How objects are grasped: the interplay between affordances and end-goals. *PLoS ONE* 6:e25203. doi: 10.1371/journal.pone.0025203

Scholz, J. P., and Schöner, G. (1999). The uncontrolled manifold concept: identifying control variables for a functional task. *Exp. Brain Res.* 126, 289–306. doi: 10.1007/s002210050738

Shadmehr, R., and Krakauer, J. W. (2008). A computational neuroanatomy for motor control. *Exp. Brain Res.* 185, 359–381. doi: 10.1007/s00221-008-1280-5

Stoianov, I., Genovesio, A., and Pezzulo, G. (2016). Prefrontal goal-codes emerge as latent states in probabilistic value learning. *J. Cogn. Neurosci.* 28, 140–157. doi: 10.1162/jocn_a_00886

Vesper, C., Butterfill, S., Knoblich, G., and Sebanz, N. (2010). A minimal architecture for joint action. *Neural Netw.* 23, 998–1003. doi: 10.1016/j.neunet.2010.06.002

Wolpert, D. M., Doya, K., and 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

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX

A.1. Relations between Coarticulation (or Assimilation) and Signaling (or Dissimilating) during Social Interactions

While we have emphasized the automaticity of coarticulation, it can also be used *strategically* in social contexts; for example, to lower (or raise) the co-actor's uncertainty about our plans—that is, to help him or her understand an actor's own distal intentions, or to feint; or even (in principle) to smoothly combine one's own actions with those of co-actors (Gonzalez et al., 2011). To illustrate how it is possible to use coarticulation strategically, we denote with $p(\mathbf{x}|m_i)$ the sequence of the states associated to the motor primitive m_i computed using the coarticulation distribution $p(\mathbf{x}_t|m_i)$. If a performer agent wants to facilitate the perceiver's action recognition process, (s)he can compute the weights $\mathbf{w}_i(t)$ so that they minimize the following equation:

$$\mathbf{w}_i(t) = argmin_{w(t)} \left[KL\left[p_i^{coa}(\mathbf{w}(t)), p_i\right] + \lambda S\left(\theta - p_i^{simulated}\right)\right] \tag{A1}$$

where:

- $KL(\cdot,\cdot)$ is the Kullback-Leibler divergence between the coarticulation distribution with the set of weights $\mathbf{w}$ and the distribution with no coarticulation;
- λ is the amount of coarticulation in the given action;
- $p_i^{simulated}$ is an estimation of the perceiving agent's posterior probability correctly recognizing the model m_i (under the assumption that performer and perceiver share the same set of internal models);
- θ is the (experimental) threshold that the agent uses during model recognition;
- S is the logistic function.

The KL term considers the cost of coarticulation, where cost can be associated to biomechanical factors, effort, and other forms of costs (e.g., cognitive costs associated to planning and executing non-familiar or non-habitual movements). The λ term permits modulation of the amount of coarticulation ($\lambda = 0$ means no coarticulation). By minimizing the above quantity, the performer agent essentially disambiguates the coarticulated action from possible alternatives, thus permitting an observer agent to infer his distal intention at early stages.

This latter example shows how it is possible to use coarticulation to *signal* one's own intentions (e.g., make them "readable"), or conversely to feint another intention, analogous to other sensorimotor communication dynamics during social interactions (Vesper et al., 2010; Pezzulo, 2011; Pezzulo and Dindo, 2013; Pezzulo et al., 2013a; Sacheli et al., 2013; Candidi et al., 2015). Indeed, in our formulation coarticulation and signaling are not just similar but stem from a consistent computational approach. Indeed, the distribution defined in Equation (2) is the dual of the *signaling* distribution defined in Pezzulo et al. (2013a), and which can be used to *dissimilate* between the current action having been performed and alternative actions, with the aim to facilitate the perceiver's agent recognition of the proximal goal.

Defining a function:

$$p^{comm}(\mathbf{x}_t|m_i;\mathbf{w}) \propto w_i \cdot p(\mathbf{x}_t|m_i) \prod_{k\in Dissim} (1 - w_k \cdot p(\mathbf{x}_t|m_k)/p_k^{max}) \cdot \prod_{j\in Assim} (w_j \cdot p(\mathbf{x}_t|m_j)) \tag{A2}$$

where p_k^{max} is the maximum value for the distribution $p(\mathbf{x}_t|m_k)$, *Dissim* is the set of motor models to be dissimulated and *Assim* is the set of motor models to be coarticulated.

In short, one can use the same equation to flexibly combine or interleave assimilation and dissimilation of actions, see **Figure 2**. As we have shown here, one can assimilate two consecutive actions, and this would correspond to coarticulation. However, one can also assimilate two simultaneous actions, and this would correspond to a feint, in that it would render the observer's action recognition process more difficult. Finally, dissimilating one's current action from the alternatives would amount to signaling (and helpful for the observer agent), while dissimilating two consecutive actions in an action sequence would amount to feinting own's own distal intention. This equation can thus be used to derive formal descriptions of various strategies to help or hinder during social interactions, which can be helpful for the (trial-by-trial, model-based) analysis of human data (Candidi et al., 2015).