Implicit Communication in Human-Robot Collaborative Transport

Elvin Yang
University of Michigan
Ann Arbor, MI, USA
eyy@umich.edu

Christoforos Mavrogiannis
University of Michigan
Ann Arbor, MI, USA
cmavro@umich.edu

***Abstract*—We focus on human-robot collaborative transport, in which a robot and a user collaboratively move an object to a goal pose. In the absence of explicit communication, this problem is challenging because it demands tight *implicit* coordination between two heterogeneous agents, who have very different sensing, actuation, and reasoning capabilities. Our key insight is that the two agents can coordinate fluently by encoding subtle, communicative signals into actions that affect the state of the transported object. To this end, we design an inference mechanism that probabilistically maps observations of joint actions executed by the two agents to a set of joint strategies of workspace traversal. Based on this mechanism, we define a cost representing the human's uncertainty over the unfolding traversal strategy and introduce it into a model predictive controller that balances between uncertainty minimization and efficiency maximization. We deploy our framework on a mobile manipulator (Hello Robot Stretch) and evaluate it in a within-subjects lab study ($N = 24$). We show that our framework enables greater team performance and empowers the robot to be perceived as a significantly more fluent and competent partner compared to baselines lacking a communicative mechanism.**

***Index Terms*—Human-robot collaboration, Human-robot teams, Implicit communication**

I. Introduction

Recently, there has been vivid interest in developing physically capable robot partners that could assist humans in *context*-rich, dynamic and unstructured domains [26] like homes [28, 32] and manufacturing sites [15]. An important task in this space involves the *collaborative transport* of objects that might be too large or too heavy to be transported by a single agent. This task is especially challenging as it not only requires efficient and fluent coordination between the two heterogeneous partners but also the simultaneous satisfaction of geometric, kinematic, and physics constraints.

Humans often tackle physically demanding collaborative tasks like transport by fluently coordinating their physical movements with their partners [25] even without a concrete plan, with minimal explicit coordination. This capability relies on sophisticated mechanisms connecting perception and action. A prevalent theory from action understanding, commonly referred to as the "teleological stance", highlights that agents' actions can often be explained by an underlying goal [1, 5, 8]. This idea has inspired researchers in human-robot interaction (HRI) to develop mechanisms that communicate a robot's intended goal to an observer through its actions [6, 12]. These

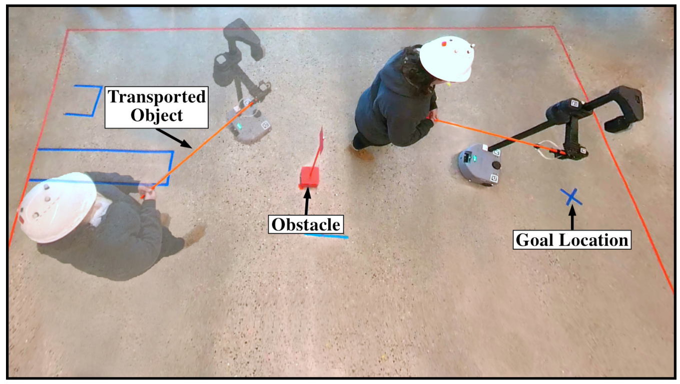

Fig. 1: Footage from our study ($N = 24$) involving the collaborative transport of an object by a user and a mobile manipulator in a workspace with an obstacle.

mechanisms have produced intent-expressive robot behavior in manipulation [6], autonomous driving [23], and social robot navigation [16].

We instantiate a task of human-robot collaborative transport, where the goal of the human-robot team is to collaboratively move an object to a goal pose while avoiding collisions with static obstacles (see Fig. 1). In this task, the user is simultaneously an observer of the robot and a *dynamic actor*, persistently influencing and being influenced by the robot while it *physically collaborates* with them. While prior work in human-robot collaborative transport has emphasized fixed leadership roles for the two agents [2, 3, 14, 18, 22, 27], we consider a *dynamic negotiation* over a joint strategy of *workspace traversal*. We contribute a control framework that leverages *implicit communication* [12] through actions influencing the state of the transported object to enable the robot to negotiate an efficient traversal with its human partner. We demonstrate our framework on a mobile manipulator and evaluate it in a lab study ($N = 24$) involving the collaborative transport of an object in a workspace with an obstacle obstruction. In our full paper [31], we perform a full analysis to show that our framework outperforms baselines lacking a communicative mechanism in terms of task completion and human impressions. Additionally, we include videos from the study (https://youtu.be/0NTDrobSifg) and code and data that could help the community iterate on our work (https://github.com/fluentrobotics/icmpc_collab_transport).

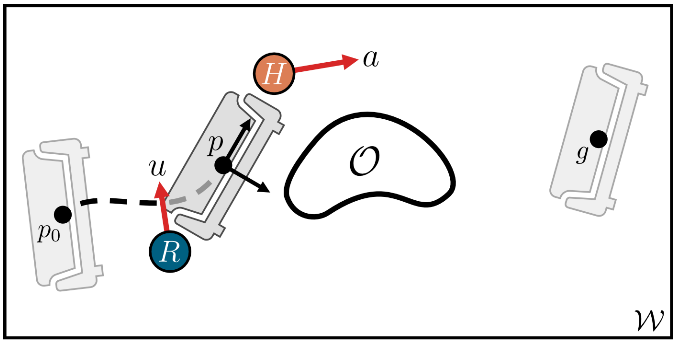

Fig. 2: A human (H) and a robot (R) collaboratively move an object from an initial pose p_0 to a final pose g in a workspace $\mathcal{W}$. An obstacle $\mathcal{O}$ stands in their way. To avoid collisions with $\mathcal{O}$ and reach g, they have to coordinate on a strategy of workspace traversal. In this work, we engineer implicit coordination through the velocities a and u that the human and the robot exert on the object.

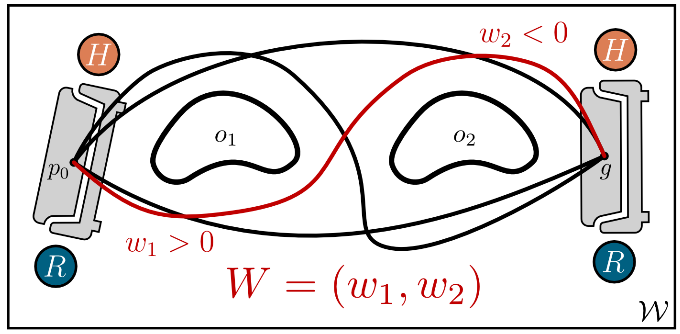

Fig. 3: Illustration of our topological abstraction for representing strategies of workspace traversal. Representing workspace traversal strategies as tuples of winding number signs, W. In this scene with two obstacles, there are four possible strategies represented as continuous curves. The red curve highlights a strategy corresponding to passing on the right of o_1 ($w_1 > 0$), and the left of o_2 ($w_2 < 0$). This representation is applicable to any number of obstacles.

II. Problem Statement

We consider a human H and a robot R collaboratively transporting an object. The robot and the human grasp the object at a fixed height; this allows us to instantiate the problem on a planar workspace $\mathcal{W} \subseteq SE(2)$. Assuming a quasistatic setting, the object's state $p \in \mathcal{W}$ evolves according to $p_{k+1} = f(p_k, a_k, u_k)$, where $a \in \mathcal{A}$, $u \in \mathcal{U}$ represent human and robot velocities, respectively, and k is a time index. The workspace includes a set of obstacle-occupied regions $\mathcal{O} \subset \mathcal{W}$. The goal of the human-robot team is to transport the object from an initial pose p_0 to a desired pose g in $\mathcal{W}$ (see Fig. 2) while avoiding collisions with $\mathcal{O}$. We assume that the two agents do not communicate explicitly (e.g., via language), but they observe the actions of one another. Our goal is to design a control policy to enable the robot to efficiently and fluently collaborate with its human partner.

III. Balancing Functional and Communicative Actions in Human-Robot Collaborative Transport

A. Formalizing Joint Strategies of Workspace Traversal

Collaborative tasks involving multiple agents working together require consensus on a *joint strategy* ψ, i.e., a qualitatively distinct way of completing the task, out of the set of all possible joint strategies, Ψ. Often, this consensus is not established *a priori*; rather, it is dynamically negotiated during execution. The abstraction of a joint strategy effectively captures critical domain knowledge at a representation level. While prior work on collaborative transport has emphasized *role* assignment across the team (i.e., whether the robot or the human are leading or following each other) [10, 19, 21], realistic, obstacle-cluttered environments present additional important challenges, such as the decision over *how to pass through* an obstacle-cluttered workspace.

In this work, we formalize the space of workspace traversal strategies using tools from homotopy theory [11]. The human-robot team is tasked with transporting an object from its initial pose p_0 to a final pose g, resulting in an object trajectory $\boldsymbol{p} : [0,1] \to \mathcal{W}$, where $\boldsymbol{p}(0) = p_0$ and $\boldsymbol{p}(1) = g$, belonging to an appropriate space of trajectories $\mathcal{P}$. Obstacles, defined as the connected components of $\mathcal{O}$, naturally partition $\mathcal{P}$ into equivalence classes Ψ, where each $\psi \in \Psi$ represents a distinct workspace traversal strategy under which the transported object can travel from p_0 to g, i.e.,

$$\begin{aligned} &\mathcal{P} = \bigcup_{\psi \in \Psi} \psi \\ &\forall \psi^i, \psi^j \in \Psi : (\psi^i \cap \psi^j \neq \emptyset) \implies (\psi^i = \psi^j) . \\ &\forall \boldsymbol{p}^i, \boldsymbol{p}^j \in \psi : \boldsymbol{p}^i \sim \boldsymbol{p}^j \end{aligned} \tag{1}$$

These classes can be identified using a notion of topological invariance. The works of Kretzschmar et al. [13], Mavrogiannis et al. [17], Vernaza et al. [29] use winding numbers to describe topological relationships between the robot and obstacles or humans navigating around it. Here, we adapt this idea to collaborative transport by enumerating the set of homotopy classes between the object trajectory and obstacles in the workspace. Specifically, for any object trajectory $\boldsymbol{p}$ embedded in a space with m obstacles $o_1, \dots, o_m$, we can define winding numbers

$$w_i = \frac{1}{2\pi} \sum_t \Delta\theta_t^i, \quad i = 1, \dots, m, \tag{2}$$

where $\Delta\theta_t^i = \angle(p_t - o_i, p_{t-1} - o_i)$ denotes an angular displacement corresponding to the transfer of the object from p_{t-1} to p_t. The sign of w_i represents the passing side between the object and the i-th obstacle, and its absolute value represents the number of times the object encircled the i-th obstacle. For a trajectory $\boldsymbol{p}$, the tuple of winding number signs

$$W = (\operatorname{sign} w_1, \dots, \operatorname{sign} w_m) \tag{3}$$

represents an equivalence class describing how the human-robot team transported the object past all obstacles in the environment. In this work, we model the space of joint strategies Ψ as set of distinct W, i.e., $|\Psi| = 2^m$.

B. Inferring Strategies of Workspace Traversal

We describe an inference mechanism that maps observations of team actions to a belief over a workspace traversal strategy.

This mechanism is agnostic to the specific definition of the strategy. At time t, we assume that the robot observes the joint action $\alpha = (a, u)$, object state p, and task context $c = (g, \mathcal{O})$. Given α, p, and c, our goal is to infer the unfolding workspace traversal strategy, ψ, i.e.,

$$\mathbb{P}(\psi \mid \alpha, p, c). \tag{4}$$

Using Bayes' rule, we can expand (4) as

$$\mathbb{P}(\psi \mid \alpha, p, c) = \frac{1}{\eta}\mathbb{P}(\alpha \mid \psi, p, c)\,\mathbb{P}(\psi \mid p, c), \tag{5}$$

where the left-hand side expression is the *posterior distribution* of the joint strategy ψ, and on the right-hand side, η is a normalizer across α, $\mathbb{P}(\alpha \mid \psi, p, c)$ is the *joint action likelihood distribution* and $\mathbb{P}(\psi \mid p, c)$ is a *prior distribution* of the joint strategy before observing the joint action. We can rewrite the joint action likelihood distribution as

$$\mathbb{P}(\alpha \mid \psi, p, c) = \mathbb{P}(a \mid \psi, p, c)\,\mathbb{P}(u \mid \psi, p, c), \tag{6}$$

since the two agents choose their actions independently.

The distribution of (4) allows the robot to represent the belief of its partner over the unfolding traversal strategy. A natural measure of uncertainty over the observer's belief regarding that strategy can be acquired by computing the information entropy of ψ, conditioned on known α, p, c:

$$H\left(\psi \mid \alpha, p, c\right) = -\sum_{\psi\in\Psi}\mathbb{P}(\psi \mid \alpha, p, c)\log\mathbb{P}(\psi \mid \alpha, p, c). \tag{7}$$

Intuitively, the higher H is, the higher the uncertainty of the user over the unfolding ψ is assumed to be.

C. Integrating Human Inferences into Robot Control

We integrate the inference mechanism of (4) into a model predictive control (MPC) algorithm by using its entropy (7) as a cost. Given the context $c = (g, \mathcal{O})$ and the object state p at time t, the goal of the MPC is to find the sequence of future robot actions $\boldsymbol{u}^*$ that minimizes a cost function J over a horizon T. At every control cycle, the MPC solves the following planning problem:

$$\begin{aligned}(u_{t:t+T})^* = \arg\min_{u_{t:t+T}} &\; J(p_{t:t+T}, u_{t:t+T})\\ s.t.\ p_{k+1} &= f(p_k, a_k, u_k)\\ a_k &\in \mathcal{A}\\ u_k &\in \mathcal{U}\end{aligned}, \tag{8}$$

We split J into a running cost J_k and a terminal cost J_T

$$\begin{aligned}J(p_{t:t+T}, u_{t:t+T}) = &\sum_{k=0}^{T}\gamma^k J_k(p_{t+k}, u_{t+k})\\ &+ J_T(p_{t+T}, u_{t+T})\end{aligned}, \tag{9}$$

where γ is a discount factor, and the terminal cost penalizes distance from the object's goal pose g:

$$J_T(p_{t+k}, u_{t+k}) = ||p_{t+k} - g||^2. \tag{10}$$

The running cost J_k is a weighted sum of two terms, i.e.,

$$\begin{aligned}J_k(p_{t+k}, u_{t+k}) = &\, w_{obs}J_{obs}(p_{t+k}, u_{t+k})\\ &+ w_{ent}J_{ent}(p_{t+k}, u_{t+k}),\end{aligned} \tag{11}$$

where

$$\begin{aligned}&J_{obs}(p_{t+k}, u_{t+k}) =\\ &\qquad \max\left(0, -\log\left(\min_{o\in\mathcal{O}}\frac{||p_{t+k} - o||}{\delta}\right)\right),\end{aligned} \tag{12}$$

is a collision avoidance cost penalizing proximity to obstacles, δ is a clearance threshold, J_{ent} is a cost proportional to the entropy defined in (7), and w_{obs}, w_{ent} are weights.

We refer to this control framework as *Implicit Communication MPC*, or **IC-MPC**. At every control cycle, IC-MPC plans a future robot trajectory that balances between functional objectives (collision avoidance, progress to goal) and communicative objectives (minimizing the partner's uncertainty over the upcoming joint strategy). The robot executes the first action u_t from the planned trajectory and then replans. This process is repeated in fixed control cycles until the task is completed.

IV. User Study

We conducted an IRB-approved, within-subjects user study (U-M HUM00254044) in which a user collaborates with a robot to jointly transport an object to a designated pose. The team operates in a workspace with area $2.8 \times 5.6\ m^2$, fitted with an overhead motion capture system that continuously streams poses and velocities of the robot and the user (via a construction-style helmet) at 120 Hz. To study the coordination of the human-robot team over a discrete decision, a single static obstacle of area $0.15 \times 0.15\ m^2$ is placed in the center of the workspace (see Fig. 1). Full experimental details and analysis can be found in [31].

A. Experiment Design

Algorithms. We compare the performance of our framework (**IC-MPC**) against two baselines:

- *Vanilla-MPC*: A purely functional ablation of IC-MPC with no uncertainty-minimizing objective ($w_{ent} = 0$).
- *VRNN* [20]: A learning-based path planner based on a Variational Recurrent Neural Network that predicts the most likely future path of the object based on human demonstrations. The robot takes actions to track path predictions as closely as possible.

Human modeling. Since our study involved an environment with a single obstacle, we set $\Psi = \{\textsc{left}, \textsc{right}\}$, corresponding respectively to $w < 0$ and $w > 0$. We instantiated the strategy inference using analytical models of the prior distribution and joint action likelihood distribution (5).

We model the prior distribution over the joint strategy as a function of the winding number w. Without loss of generality, when p_0, o, g are collinear, we consider the obstacle to have been passed when $|w| \geq \frac{1}{4}$:

$$\begin{aligned}\mathbb{P}(\textsc{left} \mid p, c) = \max\left(0, \min\left(0.5 - 2w, 1\right)\right)\\ \mathbb{P}(\textsc{right} \mid p, c) = \max\left(0, \min\left(0.5 + 2w, 1\right)\right)\end{aligned} \tag{13}$$

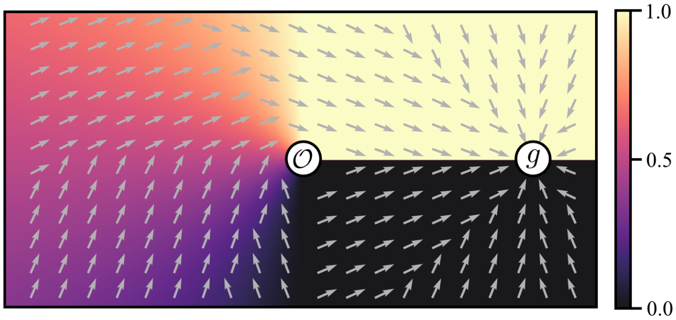

Fig. 4: Workspace traversal strategy inference. The prior distribution for $\mathbb{P}(\text{LEFT} \mid p, c)$ is shown as a colormap in the background, and the mode of the action likelihood distribution for $\mathbb{P}(a \mid \text{LEFT}, p, c)$ is shown using gray arrows in the foreground.

TABLE I: Summary of evaluation metrics. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Metric	IC-MPC	Vanilla-MPC	VRNN
Success rate (%) ↑	**98.6**	88.9	51.4
Warmth [4] ↑	3.44 (1.89)	3.12 (1.98)	2.99 (1.82)
Competence [4] ↑	**6.06 (1.86)**	5.15 (1.82)*	4.02 (1.88)***
Discomfort [4] ↓	**2.15 (1.29)**	2.86 (1.84)*	3.22 (1.49)***
Fluency [9] ↑	**5.73 (1.02)**	4.64 (1.41)**	3.67 (1.49)***

Before the obstacle is passed, the action likelihood distribution models the most likely action as the velocity that maximizes change in the winding number w. We approximate this as

$$\begin{aligned} \mathbb{P}(a \mid \text{LEFT}, p, c) &\propto \exp\left(a \cdot R\left(\frac{\pi}{3}\right)\overrightarrow{po}\right) \\ \mathbb{P}(a \mid \text{RIGHT}, p, c) &\propto \exp\left(a \cdot R\left(-\frac{\pi}{3}\right)\overrightarrow{po}\right) \end{aligned} \tag{14}$$

where $R(\cdot)$ is a 2D rotation matrix. After the obstacle is passed, the most likely action is instead in the direction of the goal. We illustrate the prior and action likelihood distributions for $\psi = \text{LEFT}$ in Fig. 4. The same model of the action likelihood distribution is used for human and robot actions. Observations of human velocities are downsampled to 10 Hz from the motion capture, and a constant velocity model is used for the human motion prediction rollouts [24].

B. Results

We report a summary of objective and subjective metrics for each algorithm in Table I. IC-MPC exhibited substantially higher *success rate* compared to both Vanilla-MPC and VRNN. Using the Friedman statistical test, we found significant effects of the robot algorithm on users' perception of *competence*, *discomfort*, and *fluency*. Post-hoc tests found that IC-MPC was judged by users as: significantly more *competent* compared to Vanilla-MPC ($p = .021$) and VRNN ($p < .001$); significantly less *discomforting* compared to Vanilla-MPC ($p = .018$) and VRNN ($p < .001$); a significantly more *fluent* collaborator compared to Vanilla-MPC ($p = .002$) and VRNN ($p < .001$).

V. Discussion

For the user study in this work, we considered coordination over a binary decision affecting a global characteristic of the trajectory followed by the human-robot team. As such, we instantiated a workspace in which the agents navigate around a single obstacle. In this workspace, we found that the computationally simple, analytical models of action likelihood and constant-velocity human motion prediction were effective to build inference and communicative mechanisms that enabled greater team performance and more positive user impressions compared to baseline methods. We note that the interaction may have been influenced by the robot's maximum speed (0.3 m/s) and payload (2 kg).

However, we also consider scenarios that admit much greater levels of complexity in the agents' movements. The two agents may need to manipulate the object along additional degrees of freedom, e.g., to lift the object over an obstacle or to pivot it through a narrow passageway. Additionally, the agents may make sudden, reactive movements, such as to balance an item on the object or to avoid dynamic obstacles. In these situations, the models we used in this study may be insufficient and impossible to reformulate analytically. To this end, in ongoing work, we aim to learn a model of human motion for collaborative transport from a dataset of human-human demonstrations [7] (Fig. 5). In contrast to human motion prediction in many existing domains of interest, human motion in collaborative transport is constrained by the physical coupling and dynamics of the two agents and the transported object. In particular, we are interested in modeling both the *motion* of the human and the *forces* and *torques* imparted on the object as a consequence of the human's motion. Thus, our model of interest takes the following form:

$$\mathbb{P}\left(x^H_{t+1:t+T}, w^H_{t+1:t+T} \;\middle|\; \begin{array}{l} x^{\text{OBJ}}_{t-H:t}, x^H_{t-H:t}, x^R_{t-H:t}, \\ w^H_{t-H:t}, w^R_{t-H:t}, \psi, c \end{array}\right), \tag{15}$$

where x^{OBJ} is the pose of the object, x^H, x^R are the kinematic states of the agents, w^H, w^R are wrenches that the two agents impart onto the object, ψ is a joint strategy, and c encapsulates task context, such as obstacles in the environment and the goal. This model is an extended implementation of the action likelihood distribution (6) used in this work, and can be incorporated into a similar posterior distribution (5) and subsequent controller, e.g., using entropy (7) or other measures of uncertainty.

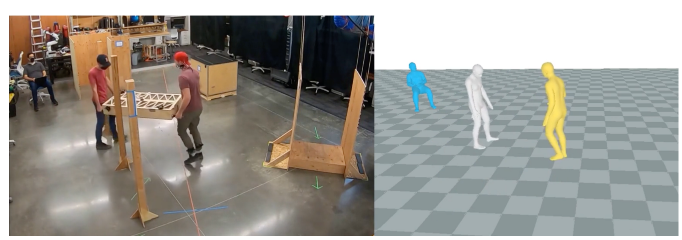

Fig. 5: Frame from the object transport dataset published by Freeman et al. [7] (left), augmented with human poses extracted using TRAM [30] (right).

References

[1] C. L. Baker, R. Saxe, and J. B. Tenenbaum. Action understanding as inverse planning. *Cognition*, 113(3): 329–349, 2009.

[2] A. Bussy, P. Gergondet, A. Kheddar, F. Keith, and A. Crosnier. Proactive behavior of a humanoid robot in a haptic transportation task with a human partner. In *Proceedings of the IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 962–967, 2012.

[3] A. Bussy, A. Kheddar, A. Crosnier, and F. Keith. Human-humanoid haptic joint object transportation case study. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Sytems (IROS)*, pages 3633–3638, 2012.

[4] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner. The robotic social attributes scale (RoSAS) development and validation. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 254–262, 2017.

[5] G. Csibra and G. Gergely. 'Obsessed with goals': Functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychologica*, 124(1):60–78, 2007.

[6] A. D. Dragan and S. Srinivasa. Integrating human observer inferences into robot motion planning. *Autonomous Robots*, 37(4):351–368, 2014.

[7] S. Freeman, S. Moss, J. L. Salmon, and M. D. Killpack. Classification of co-manipulation modus with human-human teams for future application to human-robot systems. *ACM Transactions on Human-Robot Interaction*, 13(4), 2024.

[8] G. Gergely, Z. Nádasdy, G. Csibra, and S. Bíró. Taking the intentional stance at 12 months of age. *Cognition*, 56(2):165–193, 1995.

[9] G. Hoffman. Evaluating fluency in human–robot collaboration. *IEEE Transactions on Human-Machine Systems*, 49(3):209–218, 2019.

[10] N. Jarrasse, V. Sanguineti, and E. Burdet. Slaves no longer: review on role assignment for human–robot joint motor action. *Adaptive Behavior*, 22(1):70–82, 2014.

[11] R. A. Knepper, S. S. Srinivasa, and M. T. Mason. Toward a deeper understanding of motion alternatives via an equivalence relation on local paths. *The International Journal of Robotics Research*, 31(2):167–186, 2012.

[12] R. A. Knepper, C. Mavrogiannis, J. Proft, and C. Liang. Implicit communication in a joint action. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 283–292, 2017.

[13] H. Kretzschmar, M. Spies, C. Sprunk, and W. Burgard. Socially compliant mobile robot navigation via inverse reinforcement learning. *The International Journal of Robotics Research*, 35(11):1289–1307, 2016.

[14] B. G. C. Lima, E. Ferrentino, P. Chiacchio, and M. Vento. Assistive force control in collaborative human-robot transportation. In *Proceedings of the IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 2646–2651, 2023.

[15] E. Matheson, R. Minto, E. G. Zampieri, M. Faccio, and G. Rosati. Human–robot collaboration in manufacturing applications: A review. *Robotics*, 8(4):100, 2019.

[16] C. Mavrogiannis, P. Alves-Oliveira, W. Thomason, and R. A. Knepper. Social momentum: Design and evaluation of a framework for socially competent robot navigation. *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 11(2):1–37, 2022.

[17] C. Mavrogiannis, K. Balasubramanian, S. Poddar, A. Gandra, and S. S. Srinivasa. Winding through: Crowd navigation via topological invariance. *IEEE Robotics and Automation Letters*, 8(1):121–128, 2023.

[18] E. Mielke, E. Townsend, D. Wingate, J. L. Salmon, and M. D. Killpack. Human-robot planar co-manipulation of extended objects: data-driven models and control from human-human dyads. *Frontiers in Neurorobotics*, 18, 2024.

[19] A. Mörtl, M. Lawitzky, A. Kucukyilmaz, M. Sezgin, C. Basdogan, and S. Hirche. The role of roles: Physical cooperation between humans and robots. *The International Journal of Robotics Research*, 31(13):1656–1674, 2012.

[20] E. Ng, Z. Liu, and M. Kennedy. It takes two: Learning to plan for human-robot cooperative carrying. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 7526–7532, 2023.

[21] S. Nikolaidis and J. Shah. Human-robot cross-training: Computational formulation, modeling and evaluation of a human team training strategy. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 33–40, 2013.

[22] S. Nikolaidis, D. Hsu, and S. Srinivasa. Human-robot mutual adaptation in collaborative tasks: Models and experiments. *The International Journal of Robotics Research*, 36(5-7):618–634, 2017.

[23] D. Sadigh, S. Sastry, S. A. Seshia, and A. D. Dragan. Planning for autonomous cars that leverage effects on human actions. In *Proceedings of Robotics: Science and Systems (RSS)*, 2016.

[24] C. Schöller, V. Aravantinos, F. Lay, and A. Knoll. What the constant velocity model can teach us about pedestrian motion prediction. *IEEE Robotics and Automation Letters*, 5(2):1696–1703, 2020.

[25] N. Sebanz, H. Bekkering, and G. Knoblich. Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, 10(2):70–76, 2006.

[26] M. Selvaggio, M. Cognetti, S. Nikolaidis, S. Ivaldi, and B. Siciliano. Autonomy in physical human-robot interaction: A brief survey. *IEEE Robotics and Automation Letters*, 6(4):7989–7996, 2021.

[27] J. E. Solanes, L. Gracia, P. Munoz-Benavent, J. V. Miro, M. G. Carmichael, and J. Tornero. Human–robot collaboration for safe object transportation using force

feedback. *Robotics and Autonomous Systems*, 107:196–208, 2018.

[28] T. van der Zant and L. Iocchi. Robocup@ home: Adaptive benchmarking of robot bodies and minds. In *International Conference on Social Robotics*, pages 214–225, 2011.

[29] P. Vernaza, V. Narayanan, and M. Likhachev. Efficiently Finding Optimal Winding-Constrained Loops in the Plane. In *Proceedings of Robotics: Science and Systems (RSS)*. 2013.

[30] Y. Wang, Z. Wang, L. Liu, and K. Daniilidis. Tram: Global trajectory and motion of 3d humans from in-the-wild videos. In *European Conference on Computer Vision*, pages 467–487. Springer, 2024.

[31] E. Yang and C. Mavrogiannis. Implicit communication in human-robot collaborative transport. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, page 23–33, 2025.

[32] S. Yenamandra, A. Ramachandran, M. Khanna, K. Yadav, D. S. Chaplot, G. Chhablani, A. Clegg, T. Gervet, V. Jain, R. Partsey, R. Ramrakhya, A. Szot, T.-Y. Yang, A. Edsinger, C. Kemp, B. Shah, Z. Kira, D. Batra, R. Mottaghi, Y. Bisk, and C. Paxton. The homerobot open vocab mobile manipulation challenge. In *Advances in Neural Information Processing Systems (NeurIPS): Competition Track*, 2023. URL https://aihabitat.org/challenge/2023_homerobot_ovmm/.