

Mixed-Initiative Dialog for Human-Robot Collaboration

Anonymous Author(s)

Affiliation

Address

email

1 **Abstract:** Effective robotic systems for long-horizon human-robot collaboration
2 must adapt to a wide range of human partners, whose physical behavior, will-
3 ingness to assist, and understanding of the robot's capabilities may change over
4 time. This demands a tightly coupled communication loop that grants both agents
5 the flexibility to propose, accept, or decline requests as they coordinate toward
6 completing the task effectively. We apply a Mixed-Initiative dialog paradigm to
7 Collaborative human-roBot teaming and propose **MICoBot**, a system that han-
8 dles the common scenario where both agents, using natural language, take initia-
9 tive in formulating, accepting, or rejecting proposals on who can best complete
10 different steps of a task. To handle diverse, task-directed dialog, and find suc-
11 cessful collaborative strategies that minimize human effort, MICoBot makes de-
12 cisions at three levels: (1) a meta-planner considers human dialog to formulate
13 and code a high-level collaboration strategy, (2) a planner optimally allocates the
14 remaining steps to either agent based on the robot's capabilities (measured by a
15 simulation-pretrained affordance model) and the human's estimated availability to
16 help, and (3) an action executor decides the low-level actions to perform or words
17 to say to the human. Our extensive evaluations in simulation and real-world—on a
18 physical robot with 18 unique human participants over 27 hours—demonstrate the
19 ability of our method to effectively collaborate with diverse human users, yield-
20 ing significantly improved task success and user experience than a pure LLM
21 baseline and other agent allocation models. More information on our website:
22 <https://mico-bot.github.io/>

23 **Keywords:** Human-Robot Interaction, Mixed-Initiative Grounded Dialog



Figure 1: We present MICoBot, a system for human-robot collaborative settings where both agents can initiate and carry out physical and verbal actions to negotiate how to accomplish a task together. Our system supports *both* robot-initiated (top row) *and* human-initiated (bottom row) task-directed dialog, where both agents discuss who is best suited to perform steps in a long-horizon task. The trace shows real dialog and physical interactions from our user studies (see our video and website).

24 **1 Introduction**

25 We aim to build robots that can seamlessly collaborate with humans in everyday household tasks.
26 Such collaboration is critical for deploying today’s robots with limited capabilities. To be a truly
27 effective partner, a collaborative robot must strive for task success with minimal human effort, while
28 adapting dynamically to a human user’s capabilities, preferences, and willingness to help. Beyond
29 adapting, the robot must also learn to communicate the real-time delegation of task components
30 based on which agent—human or robot—is better suited for each. Existing approaches fall short
31 of this ideal: modern AI assistants respond only to human-initiated interactions [1, 2], while prior
32 human-robot interaction (HRI) solutions often assume full control over the collaboration plan and
33 complete willingness from the human partner [3]. We argue that a truly collaborative human-robot
34 team requires a paradigm shift towards a model where both agents can take initiative to propose, bar-
35 gain, and accept or reject proposals from each other as they discuss in natural language how to best
36 complete a task. In this paper, we introduce such a collaborative system, MICoBot (Mixed-Initiative
37 Collaborative roBot), which we believe is the first to enable mixed-initiative natural language dialog
38 for real-world physical collaboration between robots and humans.

39 MICoBot enables mixed-initiative dialog to negotiate the allocation of task steps between a human
40 and a robot collaborating on a physical task, and to coordinate the physical and verbal actions needed
41 to execute the plan. We formulate this task-allocation problem as a constrained optimization where
42 the goal is to find the most suitable agent to perform each step of the task, maximizing success while
43 minimizing human effort, as well as respecting the human-initiated requests. To handle a wide
44 range of dialog, MICoBot makes optimization decisions across three levels. First, a meta-planner
45 determines the high-level strategy for collaborating with the human, incorporating human-imposed
46 constraints (such as steps they want themselves or the robot to perform), and creating reactive code
47 to generate the necessary robot actions (verbal or physical). This layer is implemented as an LLM-
48 based coder that generates adaptive planning code. Second, a planner executes the generated code
49 to determine the optimal collaboration approach, taking into account the current environment state,
50 a self-assessment of the robot’s capabilities via an affordance model trained in simulation, and a
51 dynamic evaluation of the human’s availability and willingness based on prior interactions. Finally,
52 an action executor carries out the next step of the plan, which could involve either performing a
53 manipulation action or initiating/responding to a dialog with the human.

54 Through extensive experimental evaluation, we thoroughly validate our system in both simulation
55 (with LLM-simulated humans of varying willingness and responsive moods) and the real world
56 through a user study involving 18 unique participants collaborating with a Tiago mobile manipulator
57 on three household tasks. Our approach outperforms a pure LLM baseline by **60%** in success rate,
58 while reducing the amount of human effort required compared to an all-human oracle by **60%**.
59 Additionally, our method was preferred over the pure LLM baseline by at least **80%** of participants.

60 In summary, our work’s contributions are four-fold: (1) we introduce a **new problem setting** that in-
61tegrates mixed-initiative natural language dialog with mixed-initiative human-robot interaction; (2)
62 we propose a novel **optimization framework** for task allocation, balancing human and robot effort
63 and success through a unified metric; (3) we provide a **new simulator** for collaborative household
64 tasks built on top of MiniBehavior [4] that includes LLM-controlled virtual humans and is available
65 on [our website](#); and (4) **robotic system and framework**: we develop MICoBot, a three-level hier-
66 archical solution for mixed-initiative speech2speech human-robot collaboration that flexibly adapts
67 to a wide range of real human collaborators in physically grounded, long-horizon tasks.

68 **2 Related Work**

69 **Mixed-initiative dialog** [5–7] refers to communication with freeflowing questions and answers from
70 both parties. In the NLP field, the dominant chatbot paradigm adopted by large language models
71 (LLMs) largely eschews mixed-initiative interaction: humans pose substantive questions, and the
72 chatbot primarily responds to fulfill these requests [1, 2]. Recent work has sought to make dialog

systems more goal-directed and proactive by incorporating mixed-initiative strategies—for example, persuading users to donate to charity, enhancing users’ emotional well-being [8–11], or clarifying ambiguous human requests [12–14]. However, none of these systems addressed mixed-initiative dialog in grounded, real-world collaborative scenarios involving physical manipulation tasks.

In the human-robot interaction (HRI) field, researchers have developed **human-robot collaboration systems** that interact through language but are restricted to **single-initiative dialog**. Some of these systems integrate LLMs as task planners or delegators [15–17] for tasks like real-world cooking [15] and object sorting [16]. Other systems implement a leader-follower paradigm in simulated worlds, where the leader issues natural language instructions for the follower to execute [18–21]. Single-initiative HRI systems can ask humans for clarification [22] or assistance [23–25], or inform humans of their observations [26–28]. However, by supporting only single-initiative dialog, these systems lack the capacity to adapt to the evolving nature of the human, robot, and environment—limiting their capacity to find the optimal division of labor that respects user preferences. [16].

Some works in HRI have explored **mixed-initiative collaborative systems without dialog**, only with physical actions [29–34]. In particular, Baraglia et al. [35] studied separate regimes of agent initiative (human-initiative, requesting help, or robot-initiative, proactively helping), but failed to support a natural human-robot dialog. By focusing solely on physical actions, these prior works overlook the critical role of communication in effective collaboration, thereby limiting the flexibility of the human-robot team. With MICoBot, we enable both agents to take initiative—through both physical and verbal actions—via task-grounded dialog.

Several prior works in robotics and planning have studied the problem of **human-robot optimal task allocation**, typically optimizing the time to perform a task or minimizing idle agents, posing the problem as a scheduling problem [36, 37]. Others have prioritized different objectives, such as safety [38], through the formulation of a constrained optimization problem [39]. While these solutions may achieve shorter execution times, they assume a priori known capabilities and availability of all agents, including both robots and humans. In contrast, MICoBot can adapt to the specific human’s willingness to help by estimating its availability based on previous dialog.

3 Problem Setting: Task Collaboration with Mixed-Initiative Dialog

MDP Formulation. In this paper, we study human-robot collaboration for shared manipulation problems where mixed-initiative dialog occurs. In these problems, we assume that both agents can observe the state of the world, $s \in \mathcal{S}$, and perform actions, $a \in \mathcal{A} = \mathcal{A}_p \cup \mathcal{A}_v$, comprised of a physical action space, \mathcal{A}_p , that directly affects the physical state of the environment s , (e.g., move objects, open them, etc.), and a free-form, natural language verbal action space, \mathcal{A}_v , that are directly observed by the other agent but do not change the physical state. We model the problem as a Markov Decision Process (MDP) from the robot’s point of view (see Fig. 2), where on each environment step, the robot performs some action, $a_R \in \mathcal{A}_{p,R} \cup \mathcal{A}_{v,R}$ and receives an observation $o = [I, a_{v,H}, s_{proprio}]$ consisting of an RGB-D image I , an optional verbal action from the human partner $a_{v,H}$, and the robot’s proprioceptive state $s_{proprio}$. Within each environment step, the human may perform a series of actions, $a_H \in \mathcal{A}_{p,H} \cup \mathcal{A}_{v,H}$, in its own physical and verbal action space after perceiving the world state and robot’s previous dialog, $a_{v,R}$.

Physical and Verbal Action Spaces. The physical and verbal action spaces, \mathcal{A}_p and \mathcal{A}_v , are shared between both agents. Each element of these action space are a parameterized action primitive represented by the pair, $a_{p/v} = (\omega_{p/v}, \theta_{p/v})$. ω_p is the type of the physical action primitive (open, pick-and-place, etc.) and θ_p are the corresponding parameters (e.g., what object to open or pick and where to place it). We assume that humans are fully competent in executing all steps of a collaborative household manipulation task, but may be unwilling or unavailable to perform some or

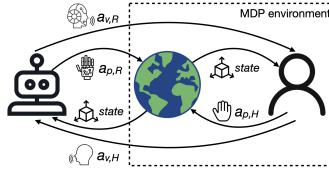


Figure 2: Our MDP Formulation for Mixed-Initiative Collaboration

123 all required actions. Their behavior can range from indifferent (never acting) to overly proactive
 124 (completing the entire task without robot involvement). In contrast, robots often have limited ma-
 125 nipulation capabilities and may be unable to execute more complex actions. ω_v is the type of the
 126 verbal action primitive (`ask_human_for_help`, `respond_to_human`, etc.) and θ_v are the corre-
 127 sponding parameters that define the necessary context of the verbal primitive (e.g., what step the
 128 robot needs help on, or can/cannot perform). While the types of verbal actions are limited, the gen-
 129 erated language based on them is freeform and open-vocabulary. MICoBot first selects an abstract
 130 verbal action from this space, then translates it into a natural language utterance to negotiate with
 131 the human—conveying its requests and the assistance it requires for successful collaboration. MI-
 132 CoBot must reason over the asymmetries in physical capabilities to devise collaboration strategies
 133 and negotiate them with verbal actions, maximizing task success while minimizing human effort.

134 **Collaborative Task Definition and Problem Statement.** We assume the collaborative task is de-
 135 fined by a task plan of length K , known to both agents and represented as a sequence of unas-
 136 signed **physical** action primitives, $[a_{p,0}, \dots, a_{p,K-1}]$, such as $[(\text{pick-and-place}(\text{box}, \text{table}), \dots,$
 137 $\text{close}(\text{box}))]$, obtained from the task instructions or off-the-shelf task planner. To complete the ma-
 138 nipulation task while minimizing human effort, the system must allocate steps of the plan between
 139 the two agents—negotiating with the human through robot-initiated dialog to suggest assignments,
 140 adapting to human preferences through human-initiated dialog, and ultimately executing its assigned
 141 physical actions. At each step t , the system must compute the best allocation of the remaining steps
 142 of the plan, $G = [g_t, \dots, g_{K-1}]$, where $\forall t, g_t \in \{H, R\}$. The optimal allocation G^* maximizes
 143 the expected task success probability while minimizing total human effort. These objectives are
 144 inherently competing: a policy focused solely on maximizing success might allocate all steps to
 145 the human (assumed to be perfectly competent); conversely, minimizing human effort alone would
 146 assign all steps to the robot, even when it may be incapable of completing certain steps. The opti-
 147 mization also incorporates constraints conveyed through the mixed-initiative dialog history, such as
 148 task allocation requests or proposed task splits. The resulting allocation G^* determines whether the
 149 robot executes the current step (R) or negotiates with the human for assistance (H).

150 4 MICoBot: Mixed-Initiative Collaborative Robot

151 Collaborative Task Allocation as Constrained Optimization.

152 In MICoBot, we formulate the step allocation problem for collaborative tasks as a constrained opti-
 153 mization, where the objective is to maximize expected task success while minimizing human effort.
 154 Constraints—such as preferences for certain steps to be done by a particular agent—are inferred
 155 through dialog with the human. To simplify the optimization and avoid a complex multi-objective
 156 formulation, we combine success probability and effort into a single cost metric for each step, re-
 157 gardless of whether it is performed by the robot or the human. Building on prior work on temporal
 158 distances in reinforcement learning [40], we use Q-functions to unify these two components. We
 159 assume each task step is executed by a multi-task policy π that operates at a fixed control frequency
 160 (e.g., once per second), performing continuous low-level control. In this low-level MDP (distinct
 161 from the high-level task MDP described in Sec. 3), we define the reward as $r = -1$ per time step
 162 until the skill completes or times out, at which point $r_{\text{termination}} = 0$. A well-trained Q-function,
 163 $Q : o_t \times a_t = (\omega_t, \theta_t) \mapsto \mathbb{R}$ with a discount factor of 1, then represents the **negative expected num-**
 164 **ber of timesteps** until skill completion from a given state. For a highly competent agent that never
 165 fail (e.g., a human), this corresponds to the average timesteps required to perform the action. For an
 166 imperfect agent that may fail, the Q-function reflects a weighted expectation over both successful
 167 and failed outcomes—where failure contributes a significant timestep penalty (timeout) weighted by
 168 its probability. We assign each agent a distinct Q-function: Q_R for the robot and Q_H for the human.
 169 These agent-specific Q-functions thus provide a unified, interpretable cost metric for comparing step
 170 allocations, jointly capturing both execution time (effort) and likelihood of success.

171 However, directly optimizing step allocation using only the Q-functions described above introduces
 172 three key limitations that diverge from realistic human-robot collaboration scenarios: (1) human

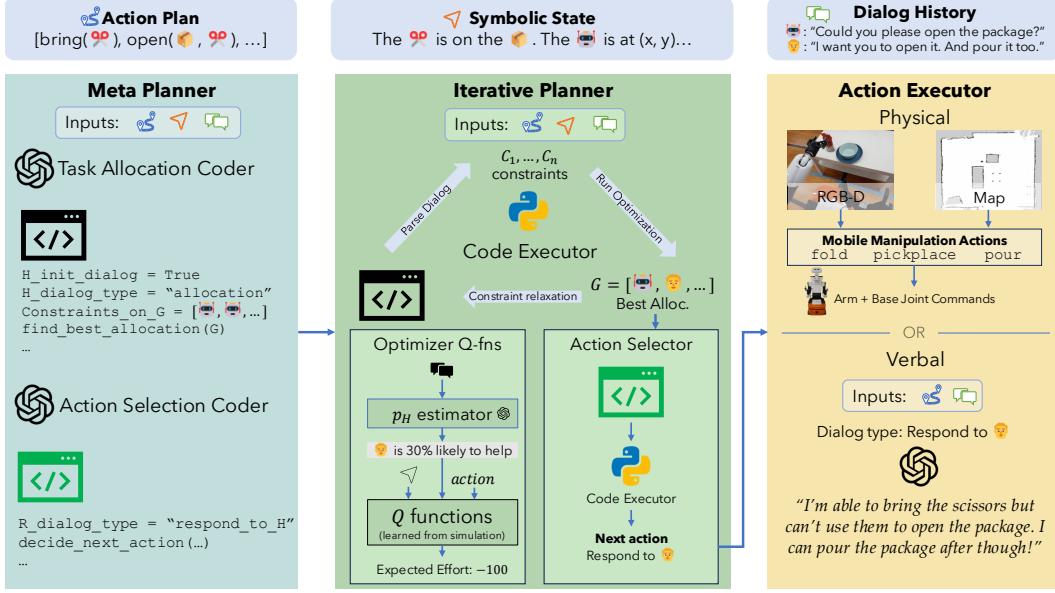


Figure 3: MICoBot consists of 3 decision-making modules: a meta-planner that outputs a strategy for task collaboration expressed through adaptive planning code, a planner that executes the code and optimizes our objective in Equation 1 to find the next primitive action to take, and the action executor that outputs the low-level physical action trajectory or verbal utterance to say to the human.

and robot effort are treated as equally, ignoring the higher value typically placed on human time and attention; (2) the human is assumed to always comply with robot-initiated requests, overlooking variability in willingness or availability; and (3) human-initiated requests or preferences are not taken into account, limiting the system’s ability to adapt to human intent. To address (1), we introduce a *human-effort factor*, α , indicating how much more valuable human effort is compared to robot effort. To address (2), we adjust the human Q-values by incorporating an inferred probability $p_{H,t}$ that represents the likelihood of the human agreeing to perform action $a_{H,t} = \omega_t(\theta_t)$ when asked. For less cooperative users, this probability lowers the expected success of the action, effectively increasing the magnitude of the negative Q-value due to potential human refusal. To address (3), we treat the optimization problem as subject to constraints, C_1, \dots, C_n , extracted from human-initiated dialog—such as explicit requests to perform specific steps themselves or to delegate them to the robot. Altogether, we propose the following objective to find the optimal task allocation G :

$$\max_{g_t, \dots, g_T} \sum_{t=1}^{T-1} \left(\mathbb{1}_{g_t=H} \cdot \frac{\alpha}{p_{H,t}} + \mathbb{1}_{g_t=R} \right) Q_{g_t}(s_t, a_t), \quad (1)$$

s.t. $C_1, \dots, C_n \leq 0$.

that minimizes expected time-to-success while prioritizing saving human effort.

4.1 MICoBot Framework

MICoBot is a three-level framework (Fig. 3) that includes 1) a meta-planner, which parses previous human dialog and generates code to optimize for task allocation and select the next action for the robot, 2) an iterative planner, which updates planning state variables, allocates and decides the next action to perform, by executing the code, and 3) an action executor, which carries out the action primitive, either through low-level physical actions or with an utterance to say to the human.

L1: Meta-planner. The meta-planner dictates the overall strategy for the lower levels to follow. Based on the most recent human dialog, the current symbolic state of the world, the task plan, and approximately 15 in-context learning (ICL) examples, it generates two pieces of code: first, **task**

195 **allocation** code to adapt the optimization computation, such as to map human dialog into additional
196 constraints, and second, **action selection** code, to determine how to choose the next action, such as
197 whether to engage in additional dialog before making further progress on the plan. The meta-planner
198 is implemented as an LLM-based (GPT-4o) coder.

199 **L2: Iterative Planner.** The iterative planner executes code generated by the meta-planner in two
200 stages. *In the first stage*, it runs the optimization routine, which enumerates all possible task al-
201 locations and selects the one that maximizes the objective in Eq. 1. To compute this, the planner
202 instantiates agent-specific Q-functions based on the current state and candidate actions (see below
203 for details on Q-function), and estimates the probability of human assistance, $p_{H,t}$, using an LLM-
204 based sentiment analysis over the prior human-robot dialog. By adjusting $p_{H,t}$, MICoBot adapts
205 to varied user sentiments by estimating the expected cost of assigning tasks to them. In the initial
206 iteration, the planner incorporates all constraints produced by the meta-planner from the mixed-
207 initiative dialog history. If no feasible allocation is found—for instance, if a human insists the robot
208 perform a step it cannot complete—the planner iteratively relaxes the most recent constraint from
209 human dialog. *In the second stage*, once the optimal allocation is determined, the planner invokes
210 meta-planner code to generate the optimal action—verbal or physical— $a = (\omega, \theta)$ to execute.

211 **L3: Action Executor.** The action executor is responsible for executing the action primitive selected
212 by the planner. For physical actions, it generates a trajectory for navigation and arm movement to
213 reach the location and manipulate the target object while avoiding obstacles. Following a similar
214 pipeline to Shah et al. [41], we use the `move_base` ROS package for path planning over a 2D
215 occupancy map, and Grounding DINO [42] to segment the target object from the scene based on the
216 natural language query in θ_t . An RGB-D camera is used to backproject segmented image pixels into
217 a 3D point cloud, from which we identify graspable or placeable points in the robot’s workspace.
218 Inverse kinematics (IK) is then used to move the arm to these points. For verbal actions, we employ
219 GPT-4o to generate natural language utterances to communicate with the human, based on both the
220 intended dialog intent (e.g., help request, split proposal) and the verbal action parameters (context
221 required for appropriate generation) from the upstream planner. Using approximately 10 in-context
222 learning (ICL) examples, the LLM produces free-form language grounded in the task context.

223 **Training Q-functions.** MICoBot’s optimization process depends on accurate approximations of
224 the Q-functions, which capture each agent’s expected effort and likelihood of success when execut-
225 ing a task step. To estimate the robot’s Q-function (Q_R), we use the OmniGibson simulator [43],
226 configured with a coarse model of the real-world task and environment (see Appendix for visual-
227 izations). In simulation, we execute action primitives that closely mirror those used by the physical
228 robot, recording both completion times and failure cases. These statistics are used to construct Q_R
229 as described earlier in this section. Since the task state is represented symbolically in both the simu-
230 lator and the real-world system, the sim-to-real gap is minimal. To estimate the human’s Q-function
231 (Q_H), we assume humans do not fail at task execution. Thus, we only require time estimates for
232 each step, which MICoBot obtains by prompting an LLM to predict how long a human would take
233 to execute action $a_t = \omega_t(\theta_t)$, plus a travel time estimate based on the human-object distances.

234 **Hierarchical Plan.** To improve communication for long-horizon task plans, MICoBot groups adja-
235 cent low-level steps into semantically meaningful abstract actions that can be discussed more suc-
236 cinctly with the human. The system only descends to a finer-grained level of detail when neces-
237 sary—during negotiation over low-level step assignments. This hierarchical approach reduces the
238 frequency and complexity of dialog, resulting in more efficient and user-friendly communication.

239 5 Evaluation

240 We evaluate MICoBot in both real-world and simulated settings. In the real world, a Tiago mobile
241 manipulator collaborates with a human user on household manipulation tasks. In simulation, we use
242 the Mini-Behavior gridworld [4] with a simulated human, allowing for larger-scale experimentation
243 and controlled comparisons across methods, particularly in relation to human behavior and dialog
244 dynamics. As suggested before, a successful robotic collaborator must complete the task efficiently

	Pour Package in Bowl n = 6		Assemble Toy Car n = 6		Pack Gift Box n = 6		Average n = 18	
	Ours	LLM	Ours	LLM	Ours	LLM	Ours	LLM
Entire Task Success Rate (%), \uparrow	50	0	67	0	67	0	61.1	0.0
% of task steps completed (\uparrow)	83	60	94	29	88	50	88.2	46.4
% of steps performed by Human	21	5	60	5	35	21	38.8	10.4
% Users Preferring ... (\uparrow)	83	17	100	0	67	33	83.3	16.7
Communicative ability (\uparrow)	3.3/5	2.3/5	4.3/5	1.3/5	2.8/5	2.3/5	3.5/5	2.0/5
“Clearly communicated to me when it couldn’t do something.”, (\uparrow)	4.3/5	2.3/5	3.7/5	1.2/5	4.2/5	2.5/5	4.1/5	2.0/5
Overall Satisfaction working w/ Robot (\uparrow)	3.7/5	2.7/5	3.5/5	1.5/5	3.5/5	2.5/5	3.6/5	2.2/5

Table 1: Comparison between our method and the LLM baseline across three real-world tasks. Ratings out of 5 are on the Likert scale. By more effective task allocation and communication, our method is able to achieve much higher task success rate and overall user satisfaction.

245 while minimizing human effort. Accordingly, our primary evaluation metric is the **success rate per unit of human effort**. We also report **subjective measures of robot behavior**, including user 246 satisfaction, preference rankings, and Likert-scale ratings.

247 **Environment.** In the real-world, we perform our experiments in a mock apartment with a kitchen 248 and living room area with commonplace furniture. In all of our tasks, the robot and human work 249 together on opposite sides of a coffee table, and the human spends most of their time on the couch, 250 where they can do their own work. The human is allowed to be as inactive or proactive as they wish 251 and to perform physical and verbal actions as defined in Section 3 (though we continue running the 252 trial if they initiate dialog beyond the scope). Each human user study consisted of two 20-30 minute 253 trials, in which they collaborated with both our method and a pure LLM baseline. The ordering of 254 the two trials was randomly determined. Trials for all methods **terminate** under any of the following 255 conditions: an irrecoverable primitive failure occurs, $4T$ steps have elapsed for a plan of length T , 256 an infeasible step is allocated to the robot twice consecutively, or the human refuses twice to perform 257 a step that the robot is incapable of executing.

258 **Baselines.** Because multiple components of our method are powered by LLMs, we compare our 259 approach to a pure LLM baseline (**LLM**) given the same information as our meta-planner: symbolic 260 state, dialog history, task plan, and α human-robot effort tradeoff factor. The LLM baseline is also 261 provided with a list of the robot’s available skills and assumes that the human always successfully 262 completes a step once they agree to perform it. The LLM baseline is prompted to produce a plan 263 allocation G that primarily optimizes for task success and secondarily minimizes human effort.

264 To control for the amount of human effort elicited in the user studies with our method, we compute 265 an additional random allocation baseline that does not involve a human participant, **RECB** (random 266 effort-controlled baseline). We denote the percentage of steps done by the human in the user trials of 267 our method as p_c . RECB randomly allocates the current step to the human with probability p_c , and 268 assumes the human always accepts the robot’s request. RECB also assumes access to oracle robot 269 primitives with 100% success rate.

270 In simulation, we additionally compare against an **RL** baseline (hierarchical task allocator + robot 271 policy; see Appendix for details), and a naive **Random** baseline, which randomly allocates either 272 agent (with probability 50%) to perform the next step.

273 **Ablations.** To measure the importance of mixed-initiative, we perform the following ablations in 274 simulation: **H-init** and **R-init**, where the human or the robot alone can initiate any dialog, respec- 275 tively. We further ablate components of MICoBot in simulation by running it **w/o P_H** (no $p_{H,t}$ 276 estimation) and **w/o Plan Hierarchy** (where our method talks to the human in low-level steps).

277 **Tasks.** We performed user studies on 3 real-world tasks (Pour Package into Bowl, Assemble Toy 278 Car, and Pack Gift Box) with 6 participants per task for a total of 18 unique human participants. 279 Each task is a long horizon sequence of 5 to 8 mobile manipulation steps. See Appendix for details.

280 **Experimental analysis.** Our experiments are designed to answer the following research questions:

281 **(1) Does our method achieve the best trade-off between task success and minimizing human 282 effort?** In our real-world user study (Table 1), MICoBot achieves a 61% task success rate, compared

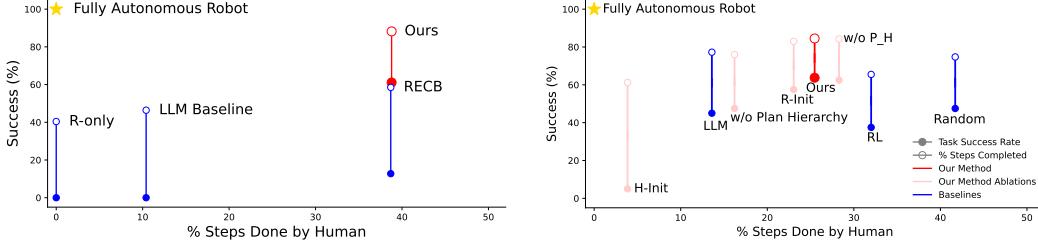


Figure 4: In both **real-world** user studies (left) and **simulation trials** with a simulated human (right), our method (red) demonstrates the best tradeoff in achieving task success (y-axis) for a given amount of human effort (x-axis) than baselines (blue) and our method’s ablations (pink).

284 to 0% for the LLM baseline, by leveraging human assistance on 38% of the steps. The LLM baseline
 285 underperformed because it prioritized minimizing human effort over task completion—requesting
 286 and receiving help in only 10% of steps, even when the robot lacked the capability to execute them.
 287 To control for the amount of human effort received, we compare our method to RECB in Figure 4.
 288 Despite RECB assuming oracle robot primitives with 100% success, our method still significantly
 289 outperforms it, demonstrating a more effective balance between success and human workload.

290 (2) How do users feel about 291 working with our system?

292 The A/B blind preference test in
 293 Table 1 shows that 83% of users
 294 preferred our method over the
 295 LLM baseline. Our method also
 296 significantly outperformed the
 297 baseline in user scores on over-
 298 all satisfaction, communicative
 299 ability, and capability in asking
 300 for a suitable amount of help (See Figure 5). In contrast, the LLM baseline consistently failed to
 301 express when it needed help and was often unwilling to reject tasks it could not complete, leading
 302 to over-promises and task failures. A representative dialog exchange—available in the Appendix
 303 and on our project website—shows MICoBot successfully persuading an initially reluctant user to
 304 perform a step the robot was incapable of executing.

305 (3) Is mixed-initiative dialog critical to our method’s performance? Figure 4 (right) shows that
 306 our full method outperforms both ablated variants that restrict dialog to single-initiative modes:
 307 robot-only initiation (R-init) and human-only initiation (H-init). H-init performs especially poorly,
 308 as it prevents the robot from requesting help for steps it cannot execute. R-init performs slightly
 309 worse than the full method because it does not allow the human to proactively initiate dialog and
 310 assist when appropriate. These results underscore the importance of mixed-initiative dialog in en-
 311 abling flexible, robust human-robot collaboration.

312 Additional experimental results and analysis (e.g. the role of $p_{H,t}$ estimation), are in the Appendix.

313 6 Conclusion

314 We proposed MICoBot, a real-world robotic collaborator that can engage in mixed-initiative dialog
 315 with humans on long-horizon mobile manipulation tasks. Our work represents the first effort to
 316 unify two previously unconnected lines of research: mixed-initiative dialog and HRI. To this end,
 317 we formulated a novel optimization function and robotic framework using mixed-initiative dialog
 318 as a rich interface for task allocation to maximize task success while minimizing human effort and
 319 complying with verbally-expressed human preferences. Real-world user studies with 18 human
 320 participants and nearly a thousand trials in simulation demonstrate the efficacy, adaptability, and
 321 user satisfaction of our method across a diverse range of human physical and verbal behavior.

322 **7 Limitations and Future Work**

323 This paper represents our initial effort on uniting mixed-initiative natural-language dialog with
324 mixed-initiative human-robot interaction. While we focused on delegating steps for long-horizon
325 manipulation tasks in a manner that maximizes task success and minimizes human effort, we be-
326 lieve this paper opens up exciting new avenues for future work. These include enabling both agents
327 learning to provide and incorporate spatial-temporal feedback to each other while performing a task,
328 share relevant task information in an imperfect-information setting, and replan and redefine a task
329 as necessary, all through mixed-initiative dialog interactions.

330 MICoBot has a number of limitations. First, it assumes a fixed plan with a predetermined ordering
331 of steps. It cannot handle cases where the human wishes to add new steps to or remove existing steps
332 from the plan dynamically, such as if the user tells the robot to “grab another cold drink while you’re
333 at the fridge before coming back to me.” Our method also cannot handle cases where a robot and
334 human wish to collaborate simultaneously on the same step in the plan, such as if the robot holds
335 a roll of tape and the human cuts from it. Furthermore, MICoBot does not support parallelization
336 where both the human and robot can work on different steps of a task simultaneously. One way to
337 address this would be to operate on plan trees, where the parent nodes are steps that must be done
338 before the child nodes, and sibling nodes can be executed by either agent in parallel.

339 Our method could be improved further by taking into account more information about the user.
340 For instance, MICoBot assumes that “effort” is based on the time necessary to perform a task until
341 completion. However, effort may also depend on the intensity of the task, how much the user enjoys
342 it, and how physically capable each user is—our method had sidestepped this issue by assuming each
343 human would expend the same amount of effort for each action primitive from some given state s .
344 Finally, there are additional ways to better predict $p_{H,t}$, such as by processing tone-of-voice and
345 observing facial expressions, that can enable the robot to produce more emotionally understanding
346 dialogue, which can potentially boost task success outcomes and increase user satisfaction.

347 **References**

- 348 [1] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal,
349 K. Slama, A. Ray, et al. Training language models to follow instructions with human feedback.
350 *Advances in neural information processing systems*, 35:27730–27744, 2022.
- 351 [2] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida,
352 J. Altenschmidt, S. Altman, S. Anadkat, et al. Gpt-4 technical report. *arXiv preprint
353 arXiv:2303.08774*, 2023.
- 354 [3] M. Selvaggio, M. Cognetti, S. Nikolaidis, S. Ivaldi, and B. Siciliano. Autonomy in physical
355 human-robot interaction: A brief survey. *IEEE Robotics and Automation Letters*, 6(4):7989–
356 7996, 2021.
- 357 [4] E. Jin, J. Hu, Z. Huang, R. Zhang, J. Wu, L. Fei-Fei, and R. Martín-Martín. Mini-behavior:
358 A procedurally generated benchmark for long-horizon decision-making in embodied ai. *arXiv
359 preprint 2310.01824*, 2023.
- 360 [5] J. R. Carbonell. Ai in cai: An artificial-intelligence approach to computer-assisted instruc-
361 tion. *IEEE Transactions on Man-Machine Systems*, 11(4):190–202, 1970. doi:10.1109/TMMS.
362 1970.299942.
- 363 [6] J. Allen, C. Guinn, and E. Horvitz. Mixed-initiative interaction. *IEEE Intelligent Systems and
364 their Applications*, 14(5):14–23, 1999. doi:10.1109/5254.796083.
- 365 [7] J. Chu-Carroll. MIMIC: An adaptive mixed initiative spoken dialogue system for infor-
366 mation queries. In *Sixth Applied Natural Language Processing Conference*, pages 97–
367 104, Seattle, Washington, USA, Apr. 2000. Association for Computational Linguistics. doi:
368 10.3115/974147.974161. URL <https://aclanthology.org/A00-1014/>.
- 369 [8] Y. Deng, W. Lei, W. Lam, and T.-S. Chua. A survey on proactive dialogue systems: Problems,
370 methods, and prospects. *arXiv preprint arXiv:2305.02750*, 2023.
- 371 [9] X. Yu, M. Chen, and Z. Yu. Prompt-based monte-carlo tree search for goal-oriented dialogue
372 policy planning, 2023. URL <https://arxiv.org/abs/2305.13660>.
- 373 [10] M. Chen, X. Yu, W. Shi, U. Awasthi, and Z. Yu. Controllable mixed-initiative dialogue gener-
374 ation through prompting. *arXiv preprint arXiv:2305.04147*, 2023.
- 375 [11] Y. Deng, W. Zhang, W. Lam, S.-K. Ng, and T.-S. Chua. Plug-and-play policy planner for large
376 language model powered dialogue agents, 2024. URL <https://arxiv.org/abs/2311.00262>.
- 377 [12] K. Qian, A. Beirami, S. Kottur, S. Shayandeh, P. Crook, A. Geramifard, Z. Yu, and C. Sankar.
378 Database search results disambiguation for task-oriented dialog systems, 2021. URL <https://arxiv.org/abs/2112.08351>.
- 379 [13] Y. Deng, L. Liao, L. Chen, H. Wang, W. Lei, and T.-S. Chua. Prompting and evaluating large
380 language models for proactive dialogues: Clarification, target-guided, and non-collaboration,
381 2023. URL <https://arxiv.org/abs/2305.13626>.
- 382 [14] M. Chen, R. Sun, S. O. Arik, and T. Pfister. Learning to clarify: Multi-turn conversations with
383 action-based contrastive self-training, 2024. URL <https://arxiv.org/abs/2406.00222>.
- 384 [15] H. Wang, K. Kedia, J. Ren, R. Abdullah, A. Bhardwaj, A. Chao, K. Y. Chen, N. Chin, P. Dan,
385 X. Fan, G. Gonzalez-Pumariega, A. Kompella, M. A. Pace, Y. Sharma, X. Sun, N. Sunkara,
386 and S. Choudhury. Mosaic: A modular system for assistive and interactive cooking, 2024.
387 URL <https://arxiv.org/abs/2402.18796>.

- 390 [16] Z. Mandi, S. Jain, and S. Song. Roco: Dialectic multi-robot collaboration with large language
 391 models, 2023. URL <https://arxiv.org/abs/2307.04738>.
- 392 [17] X. Feng, Z.-Y. Chen, Y. Qin, Y. Lin, X. Chen, Z. Liu, and J.-R. Wen. Large language model-
 393 based human-agent collaboration for complex task solving, 2024. URL <https://arxiv.org/abs/2402.12914>.
- 395 [18] A. Suhr, C. Yan, C. Schluger, S. Yu, H. Khader, M. Mouallem, I. Zhang, and Y. Artzi. Exe-
 396 cuting instructions in situated collaborative interactions, 2022. URL <https://arxiv.org/abs/1910.03655>.
- 398 [19] N. Kojima, A. Suhr, and Y. Artzi. Continual learning for grounded instruction generation by
 399 observing human following behavior, 2021. URL <https://arxiv.org/abs/2108.04812>.
- 400 [20] D. I. A. Team, J. Abramson, A. Ahuja, A. Brussee, F. Carnevale, M. Cassin, F. Fischer,
 401 P. Georgiev, A. Goldin, M. Gupta, T. Harley, F. Hill, P. C. Humphreys, A. Hung, J. Landon,
 402 T. Lillicrap, H. Merzic, A. Muldal, A. Santoro, G. Scully, T. von Glehn, G. Wayne, N. Wong,
 403 C. Yan, and R. Zhu. Creating multimodal interactive agents with imitation and self-supervised
 404 learning, 2022. URL <https://arxiv.org/abs/2112.03763>.
- 405 [21] Q. Gao, G. Thattai, S. Shakiah, X. Gao, S. Pansare, V. Sharma, G. Sukhatme, H. Shi, B. Yang,
 406 D. Zheng, L. Hu, K. Arumugam, S. Hu, M. Wen, D. Guthy, C. Chung, R. Khanna, O. Ipek,
 407 L. Ball, K. Bland, H. Rocker, Y. Rao, M. Johnston, R. Ghanadan, A. Mandal, D. H. Tur, and
 408 P. Natarajan. Alexa arena: A user-centric interactive platform for embodied ai, 2023. URL
 409 <https://arxiv.org/abs/2303.01586>.
- 410 [22] A. Z. Ren, A. Dixit, A. Bodrova, S. Singh, S. Tu, N. Brown, P. Xu, L. Takayama, F. Xia,
 411 J. Varley, Z. Xu, D. Sadigh, A. Zeng, and A. Majumdar. Robots that ask for help: Uncertainty
 412 alignment for large language model planners, 2023. URL <https://arxiv.org/abs/2307.01928>.
- 414 [23] A. Bennetot, V. Charisi, and N. Díaz-Rodríguez. Should artificial agents ask for help in human-
 415 robot collaborative problem-solving?, 2020. URL <https://arxiv.org/abs/2006.00882>.
- 416 [24] R. A. Knepper, T. Layton, J. Romanishin, and D. Rus. Ikeabot: An autonomous multi-robot
 417 coordinated furniture assembly system. In *2013 IEEE International Conference on Robotics
 418 and Automation*, pages 855–862, 2013. doi:[10.1109/ICRA.2013.6630673](https://doi.org/10.1109/ICRA.2013.6630673).
- 419 [25] M. Veloso, J. Biswas, B. Coltin, and S. Rosenthal. Cobots: robust symbiotic autonomous
 420 mobile service robots. In *Proceedings of the 24th International Conference on Artificial Intel-
 421 ligence, IJCAI’15*, page 4423–4429. AAAI Press, 2015. ISBN 9781577357384.
- 422 [26] D. L. Chen, J. Kim, and R. J. Mooney. Training a multilingual sportscaster: Using perceptual
 423 context to learn language. *Journal of Artificial Intelligence Research*, 37:397–435, 2010. URL
 424 <http://www.cs.utexas.edu/users/ai-lab?chen:jair10>.
- 425 [27] B. Mutlu, J. Forlizzi, and J. Hodgins. A storytelling robot: Modeling and evaluation of human-
 426 like gaze behavior. In *2006 6th IEEE-RAS International Conference on Humanoid Robots*,
 427 pages 518–523, 2006. doi:[10.1109/ICHR.2006.321322](https://doi.org/10.1109/ICHR.2006.321322).
- 428 [28] S. Cascianelli, G. Costante, T. A. Ciarfuglia, P. Valigi, and M. L. Fravolini. Full-gru natural
 429 language video description for service robotics applications. *IEEE robotics and automation
 430 letters*, 3(2):841–848, 2018.
- 431 [29] D. A. Few, D. J. Bruemmer, and M. C. Walton. Improved human-robot teaming through
 432 facilitated initiative. In *ROMAN 2006 - The 15th IEEE International Symposium on Robot
 433 and Human Interactive Communication*, pages 171–176, 2006. doi:[10.1109/ROMAN.2006.314413](https://doi.org/10.1109/ROMAN.2006.314413).

- 435 [30] M. Natarajan, C. Xue, S. van Waveren, K. Feigh, and M. Gombolay. Mixed-initiative human-
436 robot teaming under suboptimality with online bayesian adaptation, 2024. URL <https://arxiv.org/abs/2403.16178>.
- 438 [31] J. Bishop, J. Burgess, C. Ramos, J. B. Driggs, T. Williams, C. C. Tossell, E. Phillips, T. H.
439 Shaw, and E. J. d. Visser. Chaopt: A testbed for evaluating human-autonomy team collab-
440 oration using the video game overcooked!2. In *2020 Systems and Information Engineering
441 Design Symposium (SIEDS)*, pages 1–6, 2020. doi:[10.1109/SIEDS49339.2020.9106686](https://doi.org/10.1109/SIEDS49339.2020.9106686).
- 442 [32] A. Rosero, F. Dinh, E. J. de Visser, T. Shaw, and E. Phillips. Two many cooks: Understanding
443 dynamic human-agent team communication and perception using overcooked 2, 2021. URL
444 <https://arxiv.org/abs/2110.03071>.
- 445 [33] R. Paleja, M. Munje, K. Chang, R. Jensen, and M. Gombolay. Designs for enabling col-
446 laboration in human-machine teaming via interactive and explainable systems, 2024. URL
447 <https://arxiv.org/abs/2406.05003>.
- 448 [34] S. Jiang and R. C. Arkin. Mixed-initiative human-robot interaction: Definition, taxonomy,
449 and survey. In *2015 IEEE International Conference on Systems, Man, and Cybernetics*, pages
450 954–961, 2015. doi:[10.1109/SMC.2015.174](https://doi.org/10.1109/SMC.2015.174).
- 451 [35] J. Baraglia, M. Cakmak, Y. Nagai, R. Rao, and M. Asada. Initiative in robot assistance during
452 collaborative task execution. In *2016 11th ACM/IEEE International Conference on Human-
453 Robot Interaction (HRI)*, pages 67–74, 2016. doi:[10.1109/HRI.2016.7451735](https://doi.org/10.1109/HRI.2016.7451735).
- 454 [36] S. Vats, O. Kroemer, and M. Likhachev. Synergistic scheduling of learning and allocation of
455 tasks in human-robot teams, 2022. URL <https://arxiv.org/abs/2203.07478>.
- 456 [37] T. Yu, J. Huang, and Q. Chang. Optimizing task scheduling in human-robot collaboration
457 with deep multi-agent reinforcement learning. *Journal of Manufacturing Systems*, 60:487–
458 499, 2021. ISSN 0278-6125. doi:<https://doi.org/10.1016/j.jmsy.2021.07.015>. URL <https://www.sciencedirect.com/science/article/pii/S0278612521001527>.
- 460 [38] M. Faccio, I. Granata, and R. Minto. Task allocation model for human-robot collaboration
461 with variable cobot speed. *Journal of Intelligent Manufacturing*, 35:793–806, 2024. doi:
462 [10.1007/s10845-023-02073-9](https://doi.org/10.1007/s10845-023-02073-9). URL <https://link.springer.com/article/10.1007/s10845-023-02073-9>.
- 464 [39] S. Singh, A. Srikanthan, V. Mallampati, and H. Ravichandar. Concurrent constrained opti-
465 mization of unknown rewards for multi-robot task allocation, 2023. URL <https://arxiv.org/abs/2305.15288>.
- 467 [40] V. Myers, C. Zheng, A. Dragan, S. Levine, and B. Eysenbach. Learning temporal distances:
468 Contrastive successor features can provide a metric structure for decision-making, 2025. URL
469 <https://arxiv.org/abs/2406.17098>.
- 470 [41] R. Shah, A. Yu, Y. Zhu, Y. Zhu, and R. Martín-Martín. Bumble: Unifying reasoning and acting
471 with vision-language models for building-wide mobile manipulation. 2025.
- 472 [42] S. Liu, Z. Zeng, T. Ren, F. Li, H. Zhang, J. Yang, C. Li, J. Yang, H. Su, J. Zhu, and L. Zhang.
473 Grounding dino: Marrying dino with grounded pre-training for open-set object detection, 2023.
- 474 [43] C. Li, R. Zhang, J. Wong, C. Gokmen, S. Srivastava, R. Martín-Martín, C. Wang, G. Levine,
475 M. Lingelbach, J. Sun, M. Anvari, M. Hwang, M. Sharma, A. Aydin, D. Bansal, S. Hunter,
476 K.-Y. Kim, A. Lou, C. R. Matthews, I. Villa-Renteria, J. H. Tang, C. Tang, F. Xia, S. Savarese,
477 H. Gweon, K. Liu, J. Wu, and L. Fei-Fei. BEHAVIOR-1k: A benchmark for embodied AI
478 with 1,000 everyday activities and realistic simulation. In *6th Annual Conference on Robot
479 Learning*, 2022. URL https://openreview.net/forum?id=_8DoIe8G3t.

- 480 [44] A. Das, S. Kottur, J. M. F. Moura, S. Lee, and D. Batra. Learning cooperative visual dialog
 481 agents with deep reinforcement learning. In *Proceedings of the IEEE International Conference*
 482 on *Computer Vision (ICCV)*, Oct 2017.
- 483 [45] A. M. Shervedani, S. Li, N. Monaikul, B. Abbasi, B. D. Eugenio, and M. Zefran. Multimodal
 484 reinforcement learning for robots collaborating with humans, 2024. URL <https://arxiv.org/abs/2303.07265>.
- 486 [46] A. Padmakumar, J. Thomason, A. Shrivastava, P. Lange, A. Narayan-Chen, S. Gella, R. Pi-
 487 ramuthu, G. Tur, and D. Hakkani-Tur. Teach: Task-driven embodied agents that chat. In
 488 *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 2017–2025,
 489 2022.
- 490 [47] J. Peters, C. Waubert de Puiseau, H. Tercan, A. Gopikrishnan, G. A. Lucas de Carvalho, C. Bit-
 491 ter, and T. Meisen. Emergent language: a survey and taxonomy. *Autonomous Agents and*
 492 *Multi-Agent Systems*, 39(1), Mar. 2025. ISSN 1573-7454. doi:[10.1007/s10458-025-09691-y](https://doi.org/10.1007/s10458-025-09691-y).
 493 URL <http://dx.doi.org/10.1007/s10458-025-09691-y>.
- 494 [48] S. Kottur, J. Moura, S. Lee, and D. Batra. Natural language does not emerge ‘natur-
 495 ally’ in multi-agent dialog. In M. Palmer, R. Hwa, and S. Riedel, editors, *Proceedings of*
 496 *the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2962–
 497 2967, Copenhagen, Denmark, Sept. 2017. Association for Computational Linguistics. doi:
 498 [10.18653/v1/D17-1321](https://aclanthology.org/D17-1321). URL <https://aclanthology.org/D17-1321/>.
- 499 [49] T. Lin, M. Huh, C. Stauffer, S.-N. Lim, and P. Isola. Learning to ground multi-agent commu-
 500 nication with autoencoders, 2021. URL <https://arxiv.org/abs/2110.15349>.
- 501 [50] R. Lowe, A. Gupta, J. Foerster, D. Kiela, and J. Pineau. On the interaction between supervision
 502 and self-play in emergent communication, 2020. URL <https://arxiv.org/abs/2002.01093>.
- 504 [51] M. Woodward, C. Finn, and K. Hausman. Learning to interactively learn and assist, 2019.
 505 URL <https://arxiv.org/abs/1906.10187>.
- 506 [52] K. Lobos-Tsunekawa, A. Srinivasan, and M. Spranger. Ma-dreamer: Coordination and
 507 communication through shared imagination, 2022. URL <https://arxiv.org/abs/2204.04687>.
- 509 [53] B. Kolb, L. Lang, H. Bartsch, A. Gansekoele, R. Koopmanschap, L. Romor, D. Speck, M. Mul,
 510 and E. Bruni. Learning to request guidance in emergent communication, 2019. URL <https://arxiv.org/abs/1912.05525>.
- 512 [54] E. Jang, A. Irpan, M. Khansari, D. Kappler, F. Ebert, C. Lynch, S. Levine, and C. Finn. BC-z:
 513 Zero-shot task generalization with robotic imitation learning. In *5th Annual Conference on*
 514 *Robot Learning*, 2021. URL <https://openreview.net/forum?id=8kbp23tSGYv>.
- 515 [55] C. Lynch and P. Sermanet. Language conditioned imitation learning over unstructured data.
 516 *Robotics: Science and Systems*, 2021. URL <https://arxiv.org/abs/2005.07648>.
- 517 [56] O. Mees, L. Hermann, E. Rosete-Beas, and W. Burgard. Calvin: A benchmark for language-
 518 conditioned policy learning for long-horizon robot manipulation tasks, 2021. URL <https://arxiv.org/abs/2112.03227>.
- 520 [57] O. Mees, L. Hermann, and W. Burgard. What matters in language conditioned imitation learn-
 521 ing. *arXiv preprint arXiv:2204.06252*, 2022.
- 522 [58] L. Shao, T. Migimatsu, Q. Zhang, K. Yang, and J. Bohg. Concept2robot: Learning manip-
 523 ulation concepts from instructions and human demonstrations. In *Proceedings of Robotics:*
 524 *Science and Systems (RSS)*, 2020.

- 525 [59] S. Sodhani, A. Zhang, and J. Pineau. Multi-task reinforcement learning with context-based
 526 representations. *arXiv preprint arXiv:2102.06177*, 2021.
- 527 [60] A. Silva, N. Moorman, W. Silva, Z. Zaidi, N. Gopalan, and M. Gombolay. Lancon-learn:
 528 Learning with language to enable generalization in multi-task manipulation. In *IEEE Robotics
 529 and Automation Letters*, 2021.
- 530 [61] S. Karamcheti, M. Srivastava, P. Liang, and D. Sadigh. Lila: Language-informed latent actions.
 531 In *5th Annual Conference on Robot Learning*, 2021. URL <https://arxiv.org/pdf/2111.03205.pdf>.
- 533 [62] Y. Jiang, A. Gupta, Z. Zhang, G. Wang, Y. Dou, Y. Chen, L. Fei-Fei, A. Anandkumar, Y. Zhu,
 534 and L. Fan. Vima: General robot manipulation with multimodal prompts. *arXiv*, 2022.
- 535 [63] R. Shah, R. Martín-Martín, and Y. Zhu. Mutex: Learning unified policies from multimodal
 536 task specifications. *arXiv preprint arXiv:2309.14320*, 2023.
- 537 [64] A. Yu and R. J. Mooney. Using both demonstrations and language instructions to efficiently
 538 learn robotic tasks. *arXiv preprint arXiv:2210.04476*, 2022.
- 539 [65] W. Huang, F. Xia, T. Xiao, H. Chan, J. Liang, P. Florence, A. Zeng, J. Tompson, I. Mordatch,
 540 Y. Chebotar, P. Sermanet, N. Brown, T. Jackson, L. Luu, S. Levine, K. Hausman, and B. Ichter.
 541 Inner monologue: Embodied reasoning through planning with language models. In *arXiv
 542 preprint arXiv:2207.05608*, 2022.
- 543 [66] M. Ahn, A. Brohan, N. Brown, Y. Chebotar, O. Cortes, B. David, C. Finn, C. Fu, K. Gopalakrishnan,
 544 K. Hausman, A. Herzog, D. Ho, J. Hsu, J. Ibarz, B. Ichter, A. Irpan, E. Jang, R. J.
 545 Ruano, K. Jeffrey, S. Jesmonth, N. Joshi, R. Julian, D. Kalashnikov, Y. Kuang, K.-H. Lee,
 546 S. Levine, Y. Lu, L. Luu, C. Parada, P. Pastor, J. Quiambao, K. Rao, J. Rettinghouse, D. Reyes,
 547 P. Sermanet, N. Sievers, C. Tan, A. Toshev, V. Vanhoucke, F. Xia, T. Xiao, P. Xu, S. Xu,
 548 M. Yan, and A. Zeng. Do as I can and not as I say: Grounding language in robotic affordances.
 549 In *arXiv preprint arXiv:2204.01691*, 2022.
- 550 [67] B. Chen, F. Xia, B. Ichter, K. Rao, K. Gopalakrishnan, M. S. Ryoo, A. Stone, and D. Kappler.
 551 Open-vocabulary queryable scene representations for real world planning. In *arXiv preprint
 552 arXiv:2209.09874*, 2022.
- 553 [68] S. S. Raman, V. Cohen, D. Paulius, I. Idrees, E. Rosen, R. Mooney, and S. Tellex. Cape:
 554 Corrective actions from precondition errors using large language models. In *2nd Workshop on
 555 Language and Robot Learning: Language as Grounding*, 2023.
- 556 [69] J. Liang, W. Huang, F. Xia, P. Xu, K. Hausman, B. Ichter, P. Florence, and A. Zeng.
 557 Code as policies: Language model programs for embodied control. In *arXiv preprint
 558 arXiv:2209.07753*, 2022.
- 559 [70] C. Li, J. Liang, A. Zeng, X. Chen, K. Hausman, D. Sadigh, S. Levine, L. Fei-Fei, F. Xia, and
 560 B. Ichter. Chain of code: Reasoning with a language model-augmented code emulator. In
 561 R. Salakhutdinov, Z. Kolter, K. Heller, A. Weller, N. Oliver, J. Scarlett, and F. Berkenkamp,
 562 editors, *Proceedings of the 41st International Conference on Machine Learning*, volume 235
 563 of *Proceedings of Machine Learning Research*, pages 28259–28277. PMLR, 21–27 Jul 2024.
 564 URL <https://proceedings.mlr.press/v235/li24ar.html>.
- 565 [71] W. Huang, C. Wang, R. Zhang, Y. Li, J. Wu, and L. Fei-Fei. Voxposer: Composable 3d value
 566 maps for robotic manipulation with language models, 2023. URL <https://arxiv.org/abs/2307.05973>.
- 568 [72] L. X. Shi, B. Ichter, M. Equi, L. Ke, K. Pertsch, Q. Vuong, J. Tanner, A. Walling, H. Wang,
 569 N. Fusai, A. Li-Bell, D. Driess, L. Groom, S. Levine, and C. Finn. Hi robot: Open-ended
 570 instruction following with hierarchical vision-language-action models, 2025. URL <https://arxiv.org/abs/2502.19417>.

- 572 [73] L. X. Shi, Z. Hu, T. Z. Zhao, A. Sharma, K. Pertsch, J. Luo, S. Levine, and C. Finn. Yell at
 573 your robot: Improving on-the-fly from language corrections, 2024. URL <https://arxiv.org/abs/2403.12910>.
- 575 [74] S. Belkhale, T. Ding, T. Xiao, P. Sermanet, Q. Vuong, J. Tompson, Y. Chebotar, D. Dwibedi,
 576 and D. Sadigh. Rt-h: Action hierarchies using language, 2024. URL <https://arxiv.org/abs/2403.01823>.
- 578 [75] J.-W. Choi, H. Kim, H. Ong, Y. Yoon, M. Jang, J. Kim, et al. Reactree: Hierarchical task
 579 planning with dynamic tree expansion using llm agent nodes. 2025.
- 580 [76] X. Luo, S. Xu, and C. Liu. Obtaining hierarchy from human instructions: an llms-based
 581 approach. In *CoRL 2023 Workshop on Learning Effective Abstractions for Planning (LEAP)*,
 582 2023.
- 583 [77] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell,
 584 P. Mishkin, J. Clark, et al. Learning transferable visual models from natural language supervi-
 585 sion. *arXiv preprint arXiv:2103.00020*, 2021.
- 586 [78] X. Zhai, X. Wang, B. Mustafa, A. Steiner, D. Keysers, A. Kolesnikov, and L. Beyer. Lit:
 587 Zero-shot transfer with locked-image text tuning. *CoRR*, abs/2111.07991, 2021. URL <https://arxiv.org/abs/2111.07991>.
- 589 [79] B. Zhu, B. Lin, M. Ning, Y. Yan, J. Cui, H. Wang, Y. Pang, W. Jiang, J. Zhang, Z. Li, et al. Lan-
 590 guagebind: Extending video-language pretraining to n-modality by language-based semantic
 591 alignment. *arXiv preprint arXiv:2310.01852*, 2023.
- 592 [80] S. Nair, A. Rajeswaran, V. Kumar, C. Finn, and A. Gupta. R3m: A universal visual represen-
 593 tation for robot manipulation. *arXiv preprint arXiv:2203.12601*, 2022.
- 594 [81] M. Shridhar, L. Manuelli, and D. Fox. Cliport: What and where pathways for robotic manipu-
 595 lation. In *Proceedings of the 5th Conference on Robot Learning (CoRL)*, 2021.
- 596 [82] M. Shridhar, L. Manuelli, and D. Fox. Perceiver-actor: A multi-task transformer for robotic
 597 manipulation. *Conference on Robot Learning*, 2022.
- 598 [83] S. Nair, E. Mitchell, K. Chen, B. Ichter, S. Savarese, and C. Finn. Learning language-
 599 conditioned robot behavior from offline data and crowd-sourced annotation. In *5th Annual
 600 Conference on Robot Learning*, 2021. URL <https://arxiv.org/pdf/2109.01115.pdf>.
- 601 [84] P. Goyal, S. Niekum, and R. Mooney. Using natural language for reward shaping in reinforce-
 602 ment learning. 2019. URL <https://arxiv.org/abs/1903.02020>.
- 603 [85] P. Goyal, S. Niekum, and R. Mooney. Pixl2r: Guiding reinforcement learning using natural
 604 language by mapping pixels to rewards. 2020. URL <https://arxiv.org/abs/2007.15543>.
- 605 [86] L. Fan, G. Wang, Y. Jiang, A. Mandlekar, Y. Yang, H. Zhu, A. Tang, D.-A. Huang, Y. Zhu,
 606 and A. Anandkumar. Minedojo: Building open-ended embodied agents with internet-scale
 607 knowledge. *Advances in Neural Information Processing Systems*, 35:18343–18362, 2022.
- 608 [87] Y. J. Ma, W. Liang, V. Som, V. Kumar, A. Zhang, O. Bastani, and D. Jayaraman.
 609 Liv: Language-image representations and rewards for robotic control. *arXiv preprint
 610 arXiv:2306.00958*, 2023.
- 611 [88] Y. J. Ma, W. Liang, G. Wang, D.-A. Huang, O. Bastani, D. Jayaraman, Y. Zhu, L. Fan, and
 612 A. Anandkumar. Eureka: Human-level reward design via coding large language models, 2024.
 613 URL <https://arxiv.org/abs/2310.12931>.

- 614 [89] Y. J. Ma, W. Liang, H. Wang, S. Wang, Y. Zhu, L. Fan, O. Bastani, and D. Jayaraman.
615 Dreureka: Language model guided sim-to-real transfer. In *Robotics: Science and Systems*
616 (*RSS*), 2024.
- 617 [90] W. Yu, N. Gileadi, C. Fu, S. Kirmani, K.-H. Lee, M. G. Arenas, H.-T. L. Chiang, T. Erez,
618 L. Hasenclever, J. Humplik, B. Ichter, T. Xiao, P. Xu, A. Zeng, T. Zhang, N. Heess, D. Sadigh,
619 J. Tan, Y. Tassa, and F. Xia. Language to rewards for robotic skill synthesis, 2023. URL
620 <https://arxiv.org/abs/2306.08647>.



Figure 6: Real-world tasks from left to right: pouring package into bowl, assembling toy car, and packing gift box.

621 Appendix

622 Appendix A Real-world Task Descriptions

623 Appendix A.1 Task Plans

624 Fig. 6 depicts photos of our real-world tasks. In **Task 1: Pour Package into Bowl**, the plan includes
 625 (steps 1-3) bringing the package, scissors, and bowl from the kitchen to the coffee table, (step 4)
 626 opening the package with the scissors, and (step 5) pouring the opened package into the bowl. The
 627 robot is incapable of performing step 4 and must rely on human help. In **Task 2: Assemble Toy**
 628 **Car**, the plan includes (steps 1-3) bringing the parts tray, drill, and wheels from the shelf to the
 629 coffee table, (step 4) using the drill and wheel caps from the parts tray to put the wheels onto the
 630 chassis, (steps 5-6) finding and switching the drill bit, and (steps 7-8) screwing in the window and
 631 seats onto the car with the drill. The robot is incapable of performing steps 4, 6, 7, 8, and has a
 632 low success rate for step 5. In **Task 3: Pack Gift Box**, the plan includes (step 1) folding down the
 633 gift box flap, (steps 2-3) putting the tissue paper and toy car into the box, (steps 4-6) putting on the
 634 lid, getting the ribbons from the console table, and wrapping them around the box, and (steps 7-8)
 635 cutting a piece of tape to stick the gift bow to the top of the gift box. The robot is incapable of
 636 performing steps 4, 6, and 7, and has a low success rate for steps 2 and 5.

637 Minimal human effort required to complete the tasks ranged from just one step in Task 1 to four
 638 steps in Task 2, enabling us to test how our system compares with baselines in various regimes of
 639 dependence on human collaboration.

640 Appendix A.2 Hierarchical Plan Trees for Each Task

641 The robot assumes a high-level plan understanding with the human, and only communicates about
 642 low-level steps when necessary, such as to split up a high-level step. These are the high and low-
 643 level step breakdowns for each task, which we call the plan hierarchy. The low-level steps are listed
 644 here in skill-parameter pair format.

645 Task 1: Pour Package into Bowl (5 low-level steps)

- 646 1. Bring bowl and package to coffee table.
 - 647 (a) `pickplace(bowl, coffee_table)`
 - 648 (b) `pickplace(package, coffee_table)`
- 649 2. Open package.
 - 650 (a) `pickplace(scissors, coffee_table)`
 - 651 (b) `pick_open_place(scissors, package, coffee_table)`
- 652 3. Pour package into bowl.
 - 653 (a) `pick_pour_place(package, bowl, coffee_table)`

654 Task 2: Assemble Toy Car (8 low-level steps)

- 655 1. Bring parts to coffee table.
 - 656 (a) `pickplace(parts_tray, coffee_table)`
 - 657 (b) `pickplace(wheels, coffee_table)`
- 658 2. Assemble wheels.
 - 659 (a) `pickplace(drill, coffee_table)`
 - 660 (b) `put_on(wheels, car, drill)`
- 661 3. Switch drill bit.
 - 662 (a) `pickplace(hex_drill_bit, coffee_table)`
 - 663 (b) `switch(hex_drill_bit, drill)`
- 664 4. Assemble rest of car.
 - 665 (a) `put_on(window, car, drill)`
 - 666 (b) `put_on(seats, car, drill)`

667 Task 3: Pack Gift Box (8 low-level steps)

- 668 1. Assemble box.
 - 669 (a) `fold(box_flap)`
- 670 2. Put in gift.
 - 671 (a) `pickplace(gift_tissue_paper, box)`
 - 672 (b) `pickplace(toy_car, box)`
- 673 3. Seal the box.
 - 674 (a) `cover(box_lid, box)`
 - 675 (b) `pickplace(ribbons, coffee_table)`
 - 676 (c) `wrap(ribbons, box)`
- 677 4. Decorate the box.
 - 678 (a) `cut_put(tape, scissors, box)`
 - 679 (b) `pickplace(gift_bow, box_lid)`

680 Appendix B Dialog Excerpts from our User Studies

681 Appendix B.1 Convincing a Reluctant User to Perform a Step

682 Our system is able to flexibly handle long dialog interactions like this one on Task 1 (opening and
683 pouring the package into the bowl), where it manages to convince the human to help with opening
684 the package.

685 Robot “*Could you please open the package using the scissors on the coffee table? Thank you
686 so much!*”

687 Human “*There are no scissors on the coffee table.*”

688 Robot “*Let’s collaborate to open the package! I can place the scissors onto the coffee table,
689 and you can then use them to open the package. Thank you for your help!*”

690 Human “*I think you should bring the scissors, and then you should use the scissors to open the
691 package.*”

692 Robot “*I’m happy to place the scissors on the coffee table for you. However, I’m not able to
693 open the package with the scissors and place everything on the table. Let me know if
694 there’s anything I can do to assist.*”

695 Human “Ok, if you bring the scissors, then I will open the package with the scissors.”
 696 Robot “Great! I’ll happily place the scissors on the coffee table. Thank you for taking care of
 697 opening the package with the scissors!”
 698 Robot [brings the scissors]
 699 Human “Thanks for bringing the scissors!”
 700 Robot “Could you please open the package with scissors and place it on the coffee table?
 701 Thank you!”
 702 Human “Ok, I will do that now!”
 703 Human [opens the package]

704 **Appendix B.2 Failure Case**

705 This is an example of MICoBot unable to act on the motion-level feedback (making the tissue paper
 706 “lay flatter”) that the human wants the robot to perform. Additionally, MICoBot misunderstands the
 707 human and thinks they are talking about the ribbons step when the human is still talking about the
 708 tissue paper.

709 Human “Please pack the gift.”
 710 Robot [Puts tissue paper into the box.]
 711 Human “It would be nice if the parchment was laying flatter.”
 712 Human “The compartment is kind of squished underneath the car. If you can hold the car for
 713 me, I can fix the parchment quickly.”
 714 Robot “I’m sorry, but I can’t cover the box with the lid. Could you help with that part? Thank
 715 you so much! Could you kindly help with sealing the box? Specifically, could you cover
 716 the box with the lid? Thank you so much for your assistance!”
 717 Human “Yeah, I can, but can we fix the parchment first? If you can pick up the toy car, I can fix
 718 the parchment.”
 719 Robot “I’m delighted to hear you’ll help! Unfortunately, I won’t be able to wrap the ribbons
 720 around the box.”
 721 Human “Okay, that’s fine. We can get to that later. Let me put the lid to the box on and then we
 722 can talk about the next steps.”

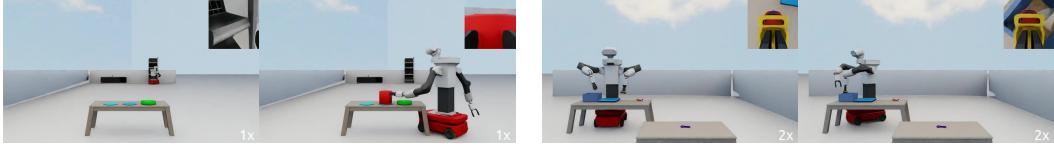
723 For the complete set of dialogs and operations in our human trials, please visit our website at <https://mico-bot.github.io>.

725 **Appendix C MICoBot Implementation Details**

726 **Appendix C.1 Robot Q-function Q_R training in OmniGibson**

727 To train Q-functions for the robot, we first create a simulated OmniGibson environment with a PAL
 728 Tiago robot and an environment that roughly matches the relative locations of the relevant furnitures
 729 and objects. We then implemented each real-world skill first in OmniGibson. Fig. 7 depicts example
 730 frames from primitives in task 1 and task 3 we ran in the OmniGibson simulator to collect sample
 731 Q-values for each skill.

732 We collected samples of the form (o, a, \mathcal{T}) , where o is the initial observation of the world, a is the
 733 skill-parameter pair (ω, θ) taken by the robot at o , and \mathcal{T} is the number of timesteps the robot takes
 734 to succeed at a from o . If the robot does not succeed in its execution, then \mathcal{T} is set to some fixed
 735 constant representing the maximum number of timesteps allowed in each skill-parameter execution.



Task 1: Pick package from shelf (left) and place on coffee table (right).

Task 3: Pick toy car from coffee table (left) and place into gift box (right).

Figure 7: Frames from primitive rollouts in OmniGibson for task 1 (left two images) and task 3 (right two images). Left and right images within each task are frames near the beginning and end, respectively, of each skill. The square image at the top right of each frame represents the robot’s camera view observation.

736 To train our Q-functions, we collect roughly 100 samples for each action a and train with inputs
 737 (o, a) and target Q-values $-\mathcal{T}$ using ℓ_2 regression with the Adam Optimizer. Our network architecture
 738 is extremely lightweight—2 linear layers with hidden size 32, and an output size of dimension 1
 739 for the Q-value.

740 Appendix C.2 Human Q-function Q_H Estimation

741 To estimate Q_H , we computed two terms. The first is the human’s stationary cost—the number of
 742 seconds it would take for the human to perform some task if the relevant items were all right in
 743 front of them. This term was copied from the output of an LLM call, which was prompted with a
 744 natural language description of the low-level step in the task, and with a URL to the toy car (for
 745 task 2). The second term is the human’s traveling time—the number of seconds it would take for
 746 the human to move from their current location to where all the objects are. This was a simple 2D
 747 euclidean distance (in meters) between the assumed human location on the couch (in the real-world
 748 user studies) and the location of the objects, divided by the average human walking speed of $1.4m/s$.
 749 We recognize this is a crude estimate of human effort, and we discuss the limitations of this in the
 750 main text.

751 Appendix C.3 Forward Dynamics Model

752 Our Q-functions rely on state and action inputs. However, computing the best task allocation in-
 753 volves considering Q-values for future steps, which depends on having knowledge of what the future
 754 state at that step will be. This involves creating a forward dynamics model so that we can estimate
 755 the future state n plan steps into the future, which can be difficult to learn accurately for continuous
 756 states. We sidestep this problem by using symbolic states for our Q-values trained in simulation,
 757 and maintaining these symbolic states during our real-world experiments. A symbolic state-based
 758 forward model is feasible to hardcode in our problem setting because we assume that each action
 759 affecting change in the world is a skill-parameter physical primitive, where the effect is quite easy
 760 to specify symbolically. For instance, the effect of `pickplace(bowl, coffee_table)` is that the
 761 bowl moves from its original furniture to the coffee table. Though this is a limitation of our method,
 762 learning a forward dynamics model is not a contribution of our work, so we leave the extension of
 763 our approach to continuous state representations to future work.

764 Appendix D Detailed Simulation Results

765 Appendix D.1 Setup

766 In simulation, we ran our method, the three baselines (RL, LLM, random), and our method’s
 767 four ablations (no $p_{H,t}$ estimation, no plan hierarchy, no R-initiative dialog, and no H-initiative
 768 dialog) on eight different settings of parameterized humans in simulation. These eight settings
 769 were a cross product of 2 dialog mood settings (positive and negative) and 4 ground-truth $\tilde{p}_{H,t} \in$
 770 $\{0.0, 0.3, 0.7, 1.0\}$ settings (following the notation introduced in Appendix G, where the \tilde{p} denotes

771 the ground truth probability while the plain p denotes our estimate). 10 trials were run for each
 772 method in each of the eight settings for the parameterized human.

773 Appendix D.2 Simulation Experiments

774 In Table 2, we show the results of our method in simulation version of our real-world Task 1. Our
 775 method performs better than baselines especially on scenarios where $\tilde{p}_{H,t}$ is low, because our method
 776 is able to take initiative in dialog, such as to propose ways to split up steps to make them more
 777 achievable with the simulated human. The averages in Table 2 are plotted in Fig. 4.

Table 2: Simulation Task 1 Performance across different $\tilde{p}_{H,t}$ Values and Language Sentiments.

Method	Metric	Human Parameters (Mood, $\tilde{p}_{H,i}$)								Avg. (%)
		Positive Mood				Negative Mood				
		0.0	0.3	0.7	1.0	0.0	0.3	0.7	1.0	
Ours	Success Rate	3/10	6/10	9/10	10/10	1/10	4/10	9/10	9/10	63.75
	Num Plan Steps Completed	3.6/5	4.2/5	4.8/5	5.0/5	3.2/5	3.8/5	4.8/5	4.5/5	84.5
	Prop. Plan Steps done by Human	0.1667	0.2381	0.3125	0.4	0.03125	0.1579	0.354	0.377	25.47
LLM Baseline	Success Rate	2/10	2/10	4/10	7/10	3/10	6/10	6/10	6/10	45
	Num Plan Steps Completed	3.4/5	3.4/5	3.7/5	4.4/5	3.6/5	4.2/5	4.0/5	4.2/5	77.25
	Prop. Plan Steps done by Human	0.0588	0.05882	0.2162	0.1591	0.1111	0.1428	0.175	0.166	13.6
Random Agent	Success Rate	2/10	5/10	6/10	7/10	2/10	3/10	6/10	7/10	47.5
	Num Plan Steps Completed	3.4/5	3.5/5	4.0/5	4.4/5	3.4/5	2.8/5	4.0/5	4.4/5	74.75
	Prop. Plan Steps done by Human	0.1176	0.4286	0.525	0.7045	0.1176	0.2143	0.525	0.7045	41.71
RL	Success Rate	0/10	1/10	4/10	10/10	0/10	1/10	4/10	10/10	37.5
	Num Plan Steps Completed	2.4/5	2.3/5	3.4/5	5.0/5	2.4/5	2.3/5	3.4/5	5.0/5	65.5
	Prop. Plan Steps done by Human	0.125	0.1739	0.4412	0.54	0.125	0.1739	0.4412	0.54	32.0
Only R Init	Success Rate	0/10	3/10	9/10	10/10	0/10	5/10	9/10	10/10	57.5
	Num Plan Steps Completed	3.0/5	3.6/5	4.8/5	5.0/5	3.0/5	4.0/5	4.8/5	5.0/5	83
	Prop. Plan Steps done by Human	0.0	0.1111	0.3542	0.4	0.0	0.225	0.354	0.4	23.05
Only H Init	Success Rate	0/10	0/10	0/10	0/10	2/10	0/10	0/10	2/10	5.0
	Num Plan Steps Completed	3.0/5	3.0/5	3.0/5	3.0/5	3.2/5	3.0/5	3.0/5	3.3/5	61.25
	Prop. Plan Steps done by Human	0.0	0.0	0.0	0.0	0.03/0.0	0.1875	0.0	0.0	0.1212
Ours w/o p_help	Success Rate	3/10	5/10	9/10	10/10	2/10	3/10	9/10	9/10	62.5
	Num Plan Steps Completed	3.6/5	4.0/5	4.8/5	5.0/5	3.4/5	3.4/5	4.8/5	4.7/5	84.25
	Prop. Plan Steps done by Human	0.1667	0.3	0.3333	0.38	0.1176	0.2059	0.3125	0.4468	28.29
Ours w/o Plan Hier.	Success Rate	2/10	4/10	7/10	10/10	0/10	3/10	4/10	8/10	47.5
	Num Plan Steps Completed	3.4/5	3.8/5	4.0/5	5.0/5	3.0/5	3.4/5	3.6/5	4.2/5	76
	Prop. Plan Steps done by Human	0.0588	0.1316	0.25	0.24	0.0667	0.1176	0.1944	0.2381	16.22

778 Appendix E User Study Details

779 Appendix E.1 User Instructions

780 Users were read the following instructions at the beginning of the study. (Instructions here are shown
 781 for task 2.)

- 782 Thank you so much for coming for our user study! We wanted to remind you to review the
 783 RIS before proceeding, and that you may voluntarily opt-out of the study at any time.
- 784 You are working with the robot to perform the task of assembling the toy car. You must use
 785 the hexagonal drill bit to screw in the wheels, and the phillips drill bit to screw in the seat
 786 and the window. [Demonstrate these steps to the human]. You and the robot operate on a
 787 shared understanding of the plan. [Read the 4 high-level steps of the plan tree for this task.]
- 788 Our goal is to simulate a home robot setting, where the human (you) are relatively busy
 789 with your own tasks, and once in a while you provide physical assistance and talk to the
 790 robot. So you are free to do work during each trial.
- 791 Once the robot asks you to do a step, and you accept, you must finish that step successfully.
- 792 We will perform 2 trials, each of a different method.
- 793 Both you and the robot can do a subset of the steps in the plan. You will communicate with
 794 the robot to determine who does what steps.
- 795 These are the objects you will work with during the task. I will move them now to their
 796 initial positions where they will start at the beginning of each trial. [Move objects to initial
 797 positions.]

- 798 8. For safety, I will gate-keep each of the robot's physical actions. In other words, the actions
 799 are generated by the robot itself, but they will be displayed on the laptop screen with a
 800 confirmation message, and I can either allow that physical action to be executed by the
 801 robot, or block the action from being executed if it brings the robot to an unsafe location.
- 802 9. The robot will stay on the TV side of the coffee table, while you will sit on the couch and
 803 stay on the couch side of the coffee table.
- 804 10. You are free to get up off the couch if you want to volunteer to perform steps that involve
 805 going to the sink or shelf, but you can only go when the robot is stationary and waiting on
 806 the other side of the coffee table. Steps are done in sequential order; our system doesn't
 807 support parallelization (agents working simultaneously).
- 808 11. You will be communicating to the robot through this headset. We will perform a mic-check
 809 now to make sure it can pick up your voice. [Do mic check.]
- 810 12. Now, this is what the robot will sound like when it talks to you. [play audio sample of the
 811 robot.] Try responding to it, and I will see if it can hear you.
- 812 13. The systems today can handle different kinds of dialog. (1) refusal/acceptance, (2) task
 813 allocation, such as ("Could you pour the package in the plate later?" Or: "I can pour the
 814 package onto the plate later."), (3) silence—you don't need to respond to the robot every
 815 time, and (4) a proposal to split up adjacent steps, such as "Please bring me the drill so that
 816 I can put on the wheels." You may engage in any of these types of dialog, and the robot
 817 may also engage in them when communicating to you.
- 818 14. Do you have any questions before we start? I will let you know when each trial begins and
 819 ends. Sometimes trials may end prematurely.

820 **Appendix E.2 Success Rate**

821 Task success at each step is measured by whether the end state has been achieved. For instance,
 822 a `pickplace(obj, furniture)` step in the plan is marked as successfully completed if the `obj`
 823 ends up on the `furniture` after execution. This means that primitive errors (such as a `pickplace`
 824 operation that accidentally moves the object off of the furniture as the arm is retracting) count as a
 825 failed execution.

826 **Appendix E.3 Failure Analysis**

827 In the real world, the main sources of error of our method were as follows:

- 828 1. Task 1: Cut and Pour Package into Bowl. 3 failed trials out of 6.
 - 829 • 2 primitive errors (pouring missed the bowl, package grasping pressed into the shelf
 830 and wasn't placed properly)
 - 831 • 1 perception error
- 832 2. Task 2: Assemble Toy Car. 2 failed trials out of 6. 1 failure that was rectified by human.
 - 833 • 1 dialog error: user wanted robot to perform a non-step plan outside of its capabilities,
 834 and refused when robot said it wasn't able to perform it
 - 835 • 1 primitive error: robot did not release its grasp of the drill.
 - 836 • 1 metaplanner dialog parsing error
- 837 3. Task 3: Pack Gift Box. 2 failed trials out of 6.
 - 838 • 1 primitive error: placed bow on box but bow dropped to floor as gripper retracted
 - 839 • 1 termination condition triggered: rejected robot's help request proposal 3 times in a
 840 row.

841 The main sources of error of the baseline was as follows:

- 842 1. Task 1: Cut and Pour Package into Bowl. 6 failed trials out of 6.
 843 • 4 task allocation errors: Allocated the infeasible package opening step to robot itself
 844 • 2 primitive errors: poured but missed the bowl; tipped over package while placing it
- 845 2. Task 2: Assemble Toy Car. 6 failed trials out of 6.
 846 • 2 primitive errors: Dropped the drill which fell off of the table when placing it
 847 • 2 perception errors: Unable to pick out correct place to place the object
 848 • 1 task allocation error: allocated to put on wheels itself.
 849 • 1 termination condition triggered: got in a conversational loop with the user.
- 850 3. Task 3: Pack Gift Box. 6 failed trials out of 6. 1 failure that was rectified by human.
 851 • 6 task allocation errors: 4 tried to put on the lid itself; 2 tried to cut a piece of tape
 852 itself.
 853 • 1 primitive error: robot inadvertently dropped the car onto the floor as it was trying to
 854 place it.

855 **Appendix E.4 Fault Recovery**

856 Sometimes, the code produced by the metaplanner is not executable. For fault recovery, the meta-
 857 planner is automatically re-queried up to 2 additional times to create code. If these attempts also
 858 produce non-executable code, the most recent dialog from the human is ignored for 2 further, au-
 859 tomatic metaplanner requeries. These re-queries are handled by a try-except block in the iterative
 860 planner module of MICoBot.

861 **Appendix F RL Baseline Details**

862 For our RL baseline which was evaluated in simulation, we trained a hierarchical policy where
 863 the high-level policy was a task allocator that outputted logits over two classes: 0 (Robot would
 864 perform current step), or 1 (Human would perform current step). If the logit for 0 were highest, then
 865 the image observation was passed into the low-level robot policy that decided the discrete physical
 866 action to take in the world. If the logit for 1 were highest, then the robot asked the human the correct
 867 verbal action for help on that step. Reward was sparse and only issued if all 5 steps were completed
 868 in the task, in the proper order.

869 We initially trained the RL policy on two simulated human settings: one where the human ground
 870 truth $\tilde{p}_{H,t} = 1.0$, and another where $\tilde{p}_{H,t} \sim U[0, 1]$. We were unable to obtain policies with any
 871 non-zero training returns after thousands of iterations on the latter setting, so we only reported results
 872 on the former setting, which explains why the RL policy does not perform well when $\tilde{p}_{H,t}$ is low.

873 **Appendix G Additional Experimental Investigations**

874 In addition to those discussed in Section 5, we explore the following additional experimental ques-
 875 tions.

876 **(4) How important is $p_{H,t}$ estimation at adapting to human collaborators?** A correct estimation
 877 of the true likeliness of a human to help, $\tilde{p}_{H,t}$, is critical: overestimating causes MICoBot to overly
 878 rely on human effort, potentially decreasing user satisfaction, while underestimating it lowers task
 879 success outcomes if the robot needs to rely on its low-success-rate skills instead of asking the human
 880 for help.

881 First, we examine in Fig. 8 a real-world instance of how well MICoBot can estimate the probability
 882 of the human helping on the next turn during the course of a user study. After the robot’s help
 883 request was rejected twice in a row (top 2 red horizontal stripes), the robot’s helpfulness estimate
 884 of the human plummets to 0.05. However, after the robot explains its incapacity to use scissors,
 885 the human accepts the next two help requests (in green) and the robot’s helpfulness estimate of the

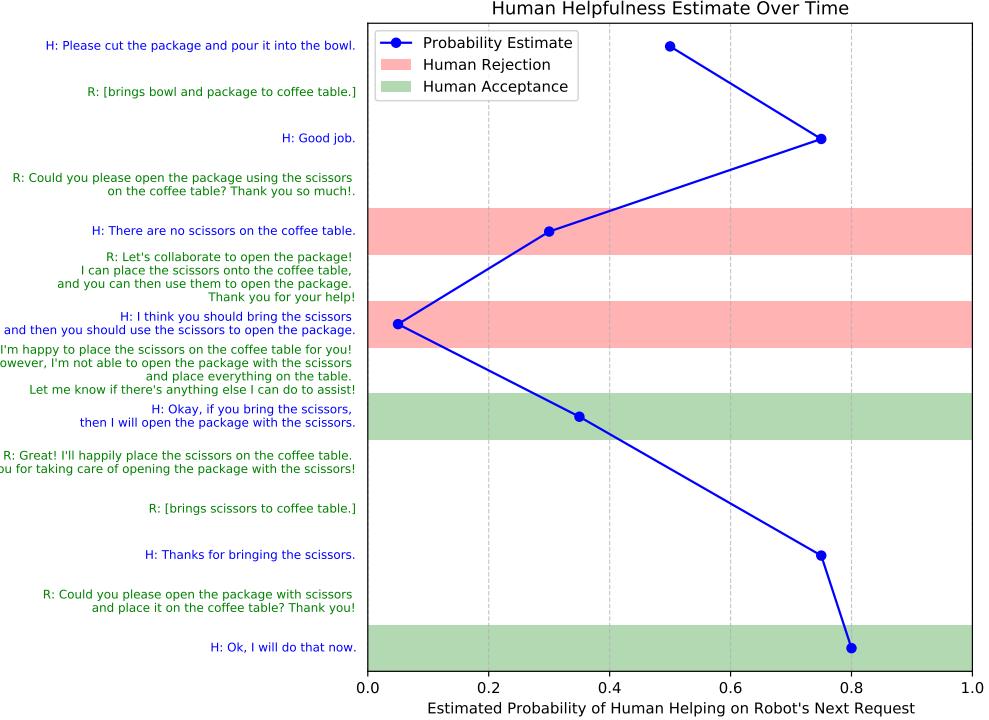


Figure 8: From a real-world user study: MICoBot’s $p_{H,t}$ estimation (blue line) reacts in real time to the human’s rejections (red), acceptances (green), and encouraging remarks. All dialog is shown as y -labels. Green text denotes robot actions/dialog, and blue text denotes human dialog. The timestep t increases from top to bottom on the y -axis.

886 human increases to 0.8. Note also that simple comments from the human, such as a “Thank you” or
 887 “good job,” also had positive effects on the estimated $p_{H,t}$, because the robot inferred that the human
 888 was in a more positive mood and hence more likely to help. This graph demonstrates that MICoBot
 889 is fairly competent at estimating a reasonable $p_{H,t}$ value when calculating the human q-values for
 890 each step in the plan.

891 To analyze the effect of a good $p_{H,t}$ estimate on task allocation, we demonstrate through a controlled
 892 toy-setting in simulation in Table 3 exactly how the optimal task allocation changes as the robot
 893 discovers more information about the human’s willingness to help. Steps that are optimally allocated
 894 to the human are shown in blue, and steps optimally allocated to the robot are shown in green.
 895 The Q-values of the selected agent in each cell are shown in parentheses. Table 3 depicts a rollout on
 896 the open and pour package into bowl (Task 1) in simulation, which has the same 5 step plan as the
 897 real-world Task 1 described in Appendix [Appendix A](#). Unlike our real-world experiments, where
 898 $\alpha = 10$, here we set $\alpha = 0.3$ so that human effort is considered around $3 \times$ *cheaper* than robot
 899 effort, which causes every task to initially get allocated to the human (at $t = 0$). Additionally, we
 900 program the human to always reject the robot’s first help request on a specific step, but to help the
 901 robot when it asks a second time.

902 As stated earlier, initially ($t = 0$) all steps are allocated to the human. However, the human rejects
 903 the initial help request from the robot, causing the $p_{H,t}$ estimate to drop to 0.25, increasing the
 904 Q-values of the human and switching the allocation of all but steps 2-3 to the robot after just two en-
 905 vironment timesteps ($t = 2$). (Recall that the robot cannot perform step 3, and due to the hierarchical
 906 structure of our plan, steps 2 and 3 are bundled together as an abstract step.) This demonstrates that
 907 having a good $p_{H,t}$ estimate is crucial to adapt to the human’s willingness to help. Since the human
 908 demonstrated initial unwillingness to help, MICoBot quickly learned to decrease its $p_{H,t}$ estimate
 909 and allocate many more of the tasks to itself by the second timestep. Had MICoBot not properly

Table 3: Computed Best Task Allocation (and Agent Q-values) During a Sim Trial on Task 1.

Env. Timestep	Step 1	Step 2	Step 3	Step 4	Step 5
$t = 0$	H (-9.6)	H (-7.2)	H (-13.2)	H (-2.4)	H (-2.4)
$t = 2$	R (-13.0)	R (-9.0)	H (-13.2)	H (-4.8)	R (-1.0)
$t = 6$	–	R (-12.0)	H (-13.2)	H (-4.8)	R (-1.0)
$t = 9$	–	–	H (-13.2)	H (-4.8)	R (-1.0)
$t = 16$	–	–	–	–	R (-3.0)

910 estimated $p_{H,t}$, it would have repeatedly asked the human for help even if the human was extremely
 911 unwilling to, leading to worse user satisfaction in working with the robot.

912 Appendix H Further Connections to Prior Work

913 Appendix H.1 Agents with Both Physical and Verbal Actions

914 MICoBot relies on a heterogeneous action space that includes interacting with the physical world
 915 *and* generating freeform dialogue to a human collaborator. Prior works have developed **policies**
 916 **with a combined physical and verbal action space** through RL [44, 45] or IL (imitation learning)
 917 [46, 20]. Research on language emergence in multiagent systems [47, 48] has also examined
 918 how cooperative agents learn to communicate through latent representations or natural language
 919 when performing simulated robotic tasks [49–53]. However, these works are typically limited to
 920 simulated domains, where action spaces, and task dynamics are highly abstracted or simplified.
 921 They often rely on limited communication protocols without integrating grounded task structure,
 922 rich human preferences, or real-world execution constraints. In contrast, MICoBot leverages an
 923 LLM to generate freeform, grounded dialogue within a shared task context, enabling fluid mixed-
 924 initiative interaction and reasoning over both verbal and physical actions in real-world scenarios.

925 Appendix H.2 Natural Language and Robotics

926 Our work sits at the broad, growing intersection of natural language and robot learning. We refer
 927 the reader to various lines of work upon which different modules of our method are based, in-
 928 cluding language-conditioned robot policies [54–64], LLMs as task planners [65–68], code-based
 929 policies [69–71], hierarchical policies [72–74] and planners [75, 76], vision-language representa-
 930 tions [77–79] for robotic control [80–82], and language-based reward shaping for RL policies [83–
 931 90].