RoboCrowd: Scaling Robot Data Collection through Crowdsourcing

Suvir Mirchandani[1], David D. Yuan[1], Kaylee Burns[1], Md Sazzad Islam[1], Tony Z. Zhao[1], Chelsea Finn[1], Dorsa Sadigh[1]

***Abstract*— In recent years, imitation learning from large-scale human demonstrations has emerged as a promising paradigm for training robot policies. However, the burden of collecting large quantities of human demonstrations is significant in terms of collection time and the need for access to expert operators. We introduce a new data collection paradigm, RoboCrowd, which distributes the workload by utilizing crowdsourcing principles and incentive design. RoboCrowd helps enable scalable data collection and facilitates more efficient learning of robot policies. We build RoboCrowd on top of ALOHA [1]—a bimanual platform that supports data collection via puppeteering—to explore the design space for crowdsourcing in-person demonstrations in a public environment. We propose three classes of incentive mechanisms to appeal to users' varying sources of motivation for interacting with the system: material rewards, intrinsic interest, and social comparison. We instantiate these incentives through tasks that include physical rewards, engaging or challenging manipulations, as well as gamification elements such as a leaderboard. We conduct a large-scale, two-week field experiment in which the platform is situated in a university café. We observe significant engagement with the system—over 200 individuals independently volunteered to provide a total of over 800 interaction episodes. Our findings validate the proposed incentives as mechanisms for shaping users' data quantity and quality. Further, we demonstrate that the crowdsourced data can serve as useful pre-training data for policies fine-tuned on expert demonstrations—boosting performance up to 20% compared to when this data is not available. These results suggest the potential for RoboCrowd to reduce the burden of robot data collection by carefully implementing crowdsourcing and incentive design principles.**

I. Introduction

With the success of pre-training large models on massive Internet-scale datasets in fields such as natural language processing and computer vision, imitation learning (IL) has become a popular paradigm for training robot policies [1]–[5]. However, modern IL algorithms continue to have significant data requirements especially as tasks increase in number and variety—on the order of hundreds to thousands of demonstrations. For example, OpenVLA [5] was trained on 970K trajectories from the Open-X Embodiment dataset [6], much of which was collected by expert human operators over the course of thousands of hours. This underscores the need for scalable methods of collecting robot data.

Prior efforts to scale up real-world data collection range from leveraging videos of human activity [7], [8] to pooling demonstration data across different institutions [6], [9], [10]. While the former approach—tapping into internet scale videos—can provide useful visual representations [11]–[13], such methods often struggle in tasks beyond pick-and-place without substantial real robot data. On the other hand, pooling datasets across many tasks and embodiments [4], [6] has amortized the cost of real-robot data collection to a degree, but expert operators are still required to collect data especially when new embodiments or tasks are added. Other works focus on how to reduce the time burden on data collectors or guide the collection strategy [14]–[16], but these methods do not address the fundamental problem that demonstrations are still solely collected by researchers or designated operators for the express purpose of training robot policies. This aspect of robot data collection drastically differs from other modalities such as text or images, where large volumes of data are *organically* produced by people in their daily activities and are *readily* available on the web. To explore ways to scale up robot data collection, we ask: *Who* can effectively collect robot data, and *how* might they be incentivized to do so?

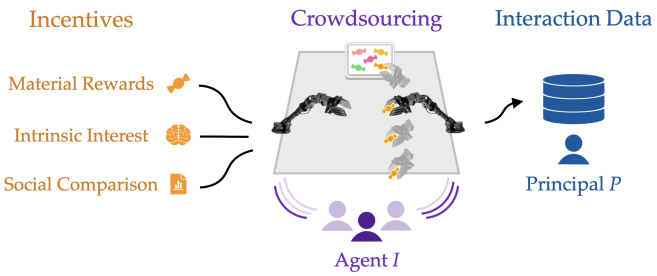


Fig. 1: **Example of incentivizing demonstrations in RoboCrowd**. The principal P consists of a robot teleoperation setup, a designer, and a scene they have designed. The scene contains tasks that an agent I (a crowd user) can attempt, guided by incentives put in place by the designer. For example, a material reward—e.g., a candy in a bin—can motivate I to produce a successful trajectory for a bin-picking task, which the designer can add to a dataset.

To tackle this problem, we look to a large body of work outside of robotics which studies strategies for incentivizing people in crowdsourced data labeling tasks [17]–[22]. The goal of these works is to align the incentives of crowdworkers with researchers' goals of labeling a given dataset—for example, *gamifying* the data labeling process [19] and aggregating data by tapping into the "wisdom of the crowd." Our key idea is to build a system that leverages similar ideas for robot data collection—i.e., *aligning human incentives to provide robot demonstration data*. However, prior strategies in human-computer interaction are designed for applications that work well with web interfaces, and applying them to robotics introduces several challenges. First, robot teleoperation traditionally requires access to physical hardware which is not readily available to crowdworkers. Second, the robot platform must be capable of performing complex tasks—in order to be engaging to users, as well as to collect useful data. At the same time, the system must be intuitive to onboard, since the vast majority of potential data providers have no teleoperation experience. Further, the system must be safe for novice users to operate.

To address these challenges, we propose RoboCrowd, a framework for incentive design in the context of crowdsourced robot

[1]Stanford University. Correspondence to suvir@cs.stanford.edu. Please see https://robocrowd.github.io for videos and appendices.

data collection. Our framework centers five key properties: public accessibility, capability, intuitiveness, safety, and gamification. Diving deeper into the incentive design problem, we incorporate three classes of incentives to appeal to users' varying sources of motivation for interacting with the system. These include *material rewards* (i.e., physical rewards from tasks), *intrinsic interest* (i.e., motivation from engaging tasks), and *social comparison* (i.e., comparison to other users). To instantiate the framework, we build upon ALOHA [1]—a bimanual platform for robot teleoperation—to satisfy our need for capable hardware, and situate the system in public spaces to enable access to general users. We design hardware enhancements and a user interface to make teleoperation intuitive and safe for non-experts. Fig. 1 illustrates an example of how incentives might shape user interactions with the system into useful data: a scene contains a bin of candies, and when a user successfully acquires a candy via teleoperation, they simultaneously contribute a trajectory to a bin-picking dataset.

We deploy the system in a field experiment in which the robot is situated near a university café, where users participate in a self-guided, gamified data collection experience. We observe significant engagement with the system—over 200 individuals independently volunteered to provide a total of over 800 interaction episodes. We compile the crowdsourced interactions into a dataset and annotate each trajectory with quality scores and task labels. We additionally validate material rewards, intrinsic interest, and social comparison as incentive types for shaping user interactions with the robot—observing up to $2\times$ the amount of data collection time spent on tasks that have preferred physical rewards (when controlling for task type) and up to $4\times$ the amount of data collection time spent on tasks that are more engaging (when controlling for physical reward). We additionally observe a positive correlation between users' response to a leaderboard (social comparison mechanism) and their data quality and quantity. Finally, we analyze the usefulness of the crowdsourced data for training policies. We demonstrate that the crowdsourced data can serve as useful pre-training data when fine-tuning on expert demonstrations, boosting policy performance up to 20% compared to expert-only policies. To our knowledge, RoboCrowd is the first system to crowdsource real-robot demonstrations for imitation learning directly from a public audience—with the potential to enable a new avenue for scalable robot data collection.

II. Related Work

In this section, we provide an overview of prior work in crowdsourcing data collection and labeling in robotics and other fields.

Crowdsourcing Non-robot Data. Crowdsourcing is a well-studied technique in human-computer interaction, often used for collecting data labels from a large set of users, with a variety of applications from computer vision to natural language processing [22]–[31]. While many works utilize platforms such as Amazon Mechanical Turk [32] and Prolific [33] to pay crowdworkers for data labels, other works consider how to *incentivize* crowdworkers via other incentives beyond direct payment to gather data [20], [34]–[37]. For example, Games-with-a-Purpose (GWAPs) [19] utilize gamification—the use of game-like elements in non-game contexts [38]—to guide users to give higher quality data labels. In [19], two players try to agree on words to describe pictures without otherwise communicating—resulting in quality image label data. Our work aims to investigate how incentive design can be adapted and applied to robot data collection—specifically, in the form of demonstrations collected via teleoperation.

Distributing Robot Data Collection. Crowdsourcing has been an attractive approach for collecting data in robotics in recent years. Prior works have attempted to crowdsource robot data via remote teleoperation in simulation or via web interfaces. RoboTurk [39], [40] develops a smartphone interface to allow crowdworkers on Mechanical Turk to collect demonstrations remotely, and shows the potential of using crowdsourced data to aid policy learning. Although this method alleviates the need for crowdworkers to physically interact with robot hardware, the ability to perform precise tasks can be limited (due to issues such as lack of depth perception, occlusion, etc.). It also presents challenges in recovering from failure states in real-world scenarios. Other works have utilized crowdsourcing to guide exploration in real-world reinforcement learning [41], [42] or to collect interaction data through high-level abstractions [43]–[45], but are again limited in the range of tasks that can be collected because they do not focus on low-level trajectory demonstrations.

Several works have developed new interfaces to make robot demonstration collection more distributed. Recent works [46]–[48] design new hardware interfaces—e.g., sensorized hand-held grippers or portable motion capture systems—to allow for demonstration collection in the real-world without needing access to a physical robot. However, crowdsourcing data with these interfaces is not immediately possible since it still requires data collectors to have access to this custom hardware. [49] presents an augmented reality tablet interface to collect robot data from everyday users, though it does not immediately extend to bimanual or dynamic tasks. In this work, rather than introducing a new teleoperation interface, we leverage an existing interface (puppeteering via ALOHA [1], which enables precise bimanual manipulation at a low-cost) and choose to situate it directly in a public space to make it accessible to data collectors. To make scaling up data collection possible, we design the system so it can be used by non-experts.

III. Preliminaries

In this section, we provide an overview of the problem of designing incentives for crowdsourcing data collection and the problem of imitation learning from collected demonstrations.

Crowdsourcing via Incentive Design. Crowdsourcing systems can be modeled as repeated principal-agent interactions. We adapt the notation from [50]. A principal P desires a pool of tasks $\mathbb{T}$ to be completed by an agent (or set of agents) I with maximum quality at minimum cost. P and I each have utility functions, denoted as $J_P : A_I \times A_P \to \mathbb{R}$ and $J_I : A_I \times A_P \to \mathbb{R}$, where A_I and A_P are the action spaces of the agent and principal respectively. An *incentive* $\gamma : A_I \to A_P$ maps between agent actions and principal actions. P aims to design γ to shape I's actions in a way that maximizes J_P, noting that for any given γ, agents have utility $J_I(a_I, \gamma(a_I))$.

In the context of crowdsourcing robot data, P abstractly represents the data collection platform and its designer, and I represents a user in the presence of the platform. For the principal P, A_P encapsulates all actions that the robot can take and how the scene changes in response to robot actions. J_P corresponds to how *useful*

the data collected by I is to P towards constructing a crowdsourced dataset $\mathcal{D}$—considering how much data is collected, of what behaviors, and of what quality. In this work, we quantify these notions in several ways (number of demonstrations, length of demonstrations, human-labeled quality scores, and downstream policy learning performance). For the agent I, A_I defines I's possible actions, such as the task choice and teleoperation actions. J_I is the utility derived by the agent from intrinsic and extrinsic factors when interacting with the robot. J_I is multifaceted and can vary widely for each I. In this work, we explore different facets of utility, such as the utility derived from receiving a physical reward as an outcome for completing a task, intrinsic interest in the task itself, and motivation driven by social comparison.

An incentive γ is a mapping from A_I to A_P. This mapping is induced by a set of decisions that P makes in developing the robot's *scene context*, such as the tasks available. To illustrate, consider a scene context consisting of a robot and a bin of physical rewards (e.g., candies), as in Fig. 1. In response to an agent action a_I (e.g., teleoperating the robot to handover a reward), the principal takes an action a_P (the robot moving as directed) which results in the agent receiving the reward. The reward is factored into the agent's utility $J_I(a_I, \gamma(a_I))$.

When the existence of an incentive γ affects I's actions such that both J_I and J_P increase, the incentive is *aligned* between P and I. For example, if I prefers to receive a candy, and teleoperates the robot in order to acquire a candy, J_I increases by the value of one candy and J_P increases in that there is one more trajectory to include in the crowdsourced dataset $\mathcal{D}$. We instantiate incentives of different classes, and illustrate that they are effective mechanisms for shaping the quality and quantity of behaviors in $\mathcal{D}$. We next describe how policies can be learned from $\mathcal{D}$ via imitation learning.

Imitation Learning. Imitation learning (IL) aims to learn a policy π_θ parameterized by θ from a dataset $\mathcal{D}$ composed of expert demonstrations. Each demonstration $\xi \in \mathcal{D}$ is a sequence of observation-action transitions $\{(o_0, a_0), \ldots, (o_T, a_T)\}$. Most commonly, IL is instantiated as behavior cloning, which trains π_θ to minimize the negative log-likelihood of data, $\mathcal{L}(\theta) = -\mathbb{E}_{(o,a)\sim\mathcal{D}}[\log \pi_\theta(a|o)]$. Since human-collected demonstrations may be diverse in practice, algorithms such as Action Chunking with Transformers (ACT) [1] are designed to model different modes of behavior. We provide an overview of ACT in Appendix VII. The success of this training paradigm hinges on the quality and quantity of trajectories in $\mathcal{D}$. We frame the creation of $\mathcal{D}$ through the lens of crowdsourcing and incentive design.

IV. RoboCrowd

In this work, we apply incentive design to the collection of robot demonstrations for imitation learning, and develop a system to collect robot demonstrations directly from the public. We propose three different incentive mechanisms to appeal to users' varying utility functions, and illustrate that incentives can impact the quantity and quality of data collected. Finally, we demonstrate the usefulness of the data for policy learning.

Enabling in-person crowdsourced teleoperation. While a sizeable body of work has studied crowdsourcing and incentive design in the context of data labeling, applying these ideas to robot demonstration collection introduces numerous challenges. First, members of the public lack direct access to robots. Additionally, implementing incentives appropriate for real-world robot demonstrations—e.g., physical rewards and intrinsically interesting tasks—requires hardware that is capable of versatile tasks. Finally, the vast majority of potential users lack experience teleoperating robots, so the system must be easy and safe for users to use. Given these challenges, we establish a set of desired properties for our system to enable crowdsourcing robot data in the real world.

- P1 *Publicly Accessible.* The system should be open to members of the public, including non-roboticists.
- P2 *Capable*. The hardware should be capable of performing complex manipulation skills.
- P3 *Intuitive*. The system should be intuitive to novice users with a self-guided onboarding process.
- P4 *Safe*. The system should be safe for novice users to operate.

Designing incentive mechanisms. Given a system that crowdworkers can interact with, we design incentive mechanisms to shape these interactions into useful data. We expect that crowdworkers may vary in their utility functions J_I. Some may be motivated by extrinsic rewards for trying out the system; others may be intrinsically interested in challenging themselves with certain tasks. Still others—e.g., people who are more competitive—may be motivated by social comparison [51]. We therefore design for three incentive mechanisms:

- M1 *Material Rewards*. Designing a scene with material rewards means that for some agent actions a_I, P performs actions $\gamma(a_I) \in A_P$ such that I receives a physical object. For example, I teleoperating a bin-picking task results in P performing an action which delivers the reward to I.
- M2 *Intrinsic Interest*. Designing a scene for intrinsic interest expands A_P to include engaging or challenging tasks. For example, as the result of certain agent teleoperation actions a_I, P may perform fine-grained object manipulation $\gamma(a_I)$.
- M3 *Social Comparison*. Designing a scene to enable social comparison involves a mechanism along which agents can compare themselves. For example, the action a_I of teleoperating a successful trajectory can result in a principal action $\gamma(a_I)$ which awards the agent points and increases their position on a leaderboard.

P1-4 are prerequisites to crowdsourcing; our approach to effectively implementing M1-3 involves an additional characteristic:

- P5 *Gamified.* The system should permit gamified elements—e.g., the ability to track individual users and the ability to provide physical rewards.

The rest of this section explains how we meet these desiderata through our hardware and software design.

A. Hardware Design

We select ALOHA [1], a system for bimanual teleoperation, as the base platform for our system. ALOHA consists of two "follower" arms (ViperX) that are controlled via puppeteering with two "leader" arms (WidowX). We choose to use the ALOHA platform due to its low-cost, repairability, as well as its ability for collecting data for a wide task range. Fig. 2 illustrates a set of enhancements to outfit ALOHA for public use to achieve our desired properties and enable crowdsourcing. First, we implement

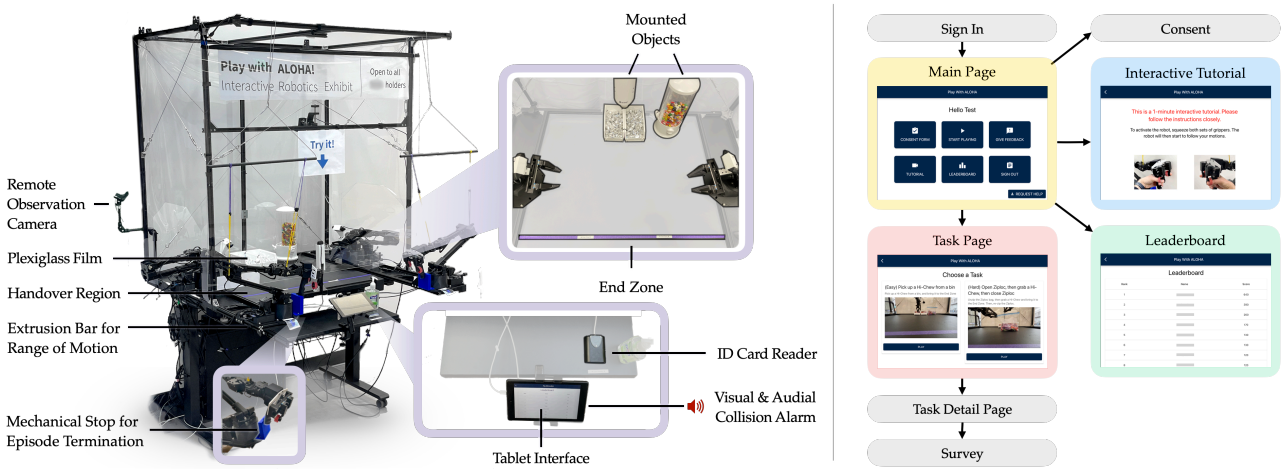


Fig. 2: **System Overview**. (*Left*) RoboCrowd uses the ALOHA robot [1], a bimanual teleoperation platform wherein users control 2 ViperX follower arms by puppeteering via 2 WidowX leader arms. Users can perform tasks in scenes put in place by the scene designer; tasks may include physical rewards that the user can bring to the End Zone and access via the Handover Region. (*Right*) Users are guided by a GUI on a tablet. Functionalities include an Interactive Tutorial to get acquainted with RoboCrowd, a Task Page to select among tasks, and a Leaderboard where users can compare their scores. For additional details, please see Appendix V.

mechanisms for user and robot safety (P4): (a) collision avoidance to prevent self-collisions, achieved via a parallel MuJoCo [52] simulator, as well as a visual-audial alarm when the robot is near collision; (b) plexiglass and vinyl film to cover all sides of the ALOHA workcell to enclose the puppet arms; (c) extended extrusion bars on the leader arms to increase the distance between users and leader arms; (d) mounting of scene props (such as bins and dispensers) to mitigate scene damage; and (e) a remote observation camera for the scene designer to periodically monitor the scene. We also include enhancements to increase the intuitiveness of the platform for members of the public (P3): (a) a tablet interface, described in the next section; (b) a mechanical stop for users to automatically terminate episodes by resting the puppet arms. To enable a gamified setup (P5), we utilize (a) an ID card reader to authenticate and track users and (b) demarcate an "End Zone" within scenes, where a user a can place physical rewards and access them via a handover region at the bottom of the plexiglass casing. Given its ability to perform versatile tasks, ALOHA satisfies our capability goal (P2). We physically situate it in a public environment (Section V) to make it accessible to crowd users (P1).

B. *Software Design*

To make operating the robot intuitive (P2) for members of the public, we implement a tablet application to complement the hardware platform and guide users through the operation process (Fig. 2; right). The interface additionally features a variety of elements of gamification (P5) that we highlight below.

Onboarding. We develop an onboarding process for new users to sign-in and receive a tutorial to familiarize themselves with the platform. In pilot studies (Appendix VI) where users were asked to use the system but were not given further verbal instructions (to mimic organic encounters that crowd users might have), users reported a desire for "instant gratification" and wished to begin to use the robot as soon as possible rather than watching a video or reading instructions. Thus, we design our onboarding process to be efficient and interactive: users begin by tapping their university ID card on a card reader, which directs them to a Sign In page to create a *user profile*. Users are then directed to complete a consent form and an interactive tutorial to learn how to puppeteer the robot (Fig. 2; right). The tutorial contains four steps and takes less than one minute to complete. We detail the stages of the interactive tutorial in the Appendix V.

Fig. 3: **Scene Setup**. Illustration of `BinScene`, `Bin+DispenserScene`, and `Bin+ZiplocScene`, and the objects relevant to our 6 tasks (`hi-chew`, `tootsie-roll`, `hershey-kiss`, `jelly-bean`, `hi-chew-bin`, `hi-chew-ziploc`).

Performing Tasks. After completing the tutorial, users can choose to enter a *Task Page* where they see videos of different tasks they can complete in the scene (Fig. 2; right). These tasks can be presented in various ways; for example, marked with *levels of difficulty* (e.g., easy versus hard). In service of P5, we use gamified verbiage and elements throughout the interface (e.g. a *Start Playing* button, and a *countdown timer* on performing tasks). Specifically for M3, we implement a point system where users receive points for completing tasks, which are tallied and visible on a *Leaderboard Page*, where users can see how their scores rank compared to other users (Fig. 2; right). We describe implementation details of the software architecture in Appendix V.

V. Experimental Setup

We utilize RoboCrowd to collect a crowdsourced dataset over a two-week period in a public university café. We instantiate three types of incentive mechanisms (M1-M3) to appeal to users' varying utility functions J_I, and design scenes in order to verify if these mechanisms can shape demonstration quantity and quality. This section details our experimental setup. We then analyze the data and discuss the results in Section VI.

Scene Design. On each day of crowdsourcing, two of six tasks are made available to users, with different pairs corresponding to different scenes (Fig. 3). `BinScene` contains bins with two types of

candies for single arm bin-picking tasks (`hi-chew` and `tootsie-roll`). `Bin+DispenserScene` contains the same bins with a single type of candy (`hershey-kiss`), as well as a cup dispenser and a jelly bean dispenser (`jelly-bean`). `Bin+ZiplocScene` contains the same bins with a single candy type (`hi-chew-bin`) as well as a closed Ziploc bag full of candies (`hi-chew-ziploc`).

Incentive Types. We select tasks to study 3 classes of incentives.

[M1] *Material Rewards.* We hypothesize that direct material rewards can influence which tasks users perform with the system. We design a simple scene context to test this (`BinScene`). There are two bins on the table, one containing Hi-Chews and the other containing Tootsie Rolls (Fig. 3). There are two bin-picking tasks available to the user on the Task Page: "pick up Hi-Chew" (`hi-chew`) and "pick up Tootsie Roll" (`tootsie-roll`). We hypothesize that users who engage with the robot will more often choose to interact with the Hi-Chew (which, in an offline survey, we find is more desired than the Tootsie Roll; see Appendix III). This incentive mechanism is an example of *extrinsic motivation*.

[M2] *Intrinsic Interest.* Users may also be *intrinsically* motivated in how they choose to interact with the system. We hypothesize that users prefer to spend time on tasks that are more qualitatively interesting and challenging. Therefore, we design `Bin+ZiplocScene` to contain a bin with Hi-Chews as well as a closed Ziploc bag with Hi-Chews inside. This scene features two available tasks: "pick up Hi-Chew from bin" and "open Ziploc, pick up Hi-Chew, close Ziploc." With the same extrinsic reward, the latter task is *significantly* more challenging, yet may be more intrinsically interesting to users. We test this effect in `Bin+DispenserScene` as well, which contains a bin with Hershey Kisses as well as a cup dispenser and a dispenser containing Jelly Beans. The tasks available to the user in this scene are "pick up Hershey Kiss from bin" (`hershey-kiss`) and "take cup from dispenser and eject Jelly Bean into the cup" (`jelly-bean`). The latter task is again significantly more challenging; but note that it does not provide greater extrinsic reward according to our offline survey (see Appendix III).

[M3] *Social Comparison.* Users may vary in how they respond to gamification mechanisms for social comparison in the interface. To test the idea that gamified elements can shape the way certain users collect data, we include a leaderboard that tallies the number of "points" users achieve by completing tasks (Fig. 2). Using quantitative measurements to compare players, including via leaderboards, is a common method for provoking competition [53]. We hypothesize that users who choose to look at the leaderboard may give a higher quantity of data, stemming from social comparison as an incentive mechanism.

Data Annotation Pipeline. User interaction data are a mixture of task-relevant data, tutorial interactions, and "play" data. We manually annotate all interactions by whether the user was engaging in free play or task-relevant behavior, as well as quality scores on a scale of 0 (play data) to 3 (highest quality task data). We define these quality labels based on how smooth the user's motions are, whether there is retrying behavior or extraneous movements, etc. Importantly, each interaction episode may include data relevant to different tasks and of various qualities, so we annotate with these labels at *every transition* per trajectory. For more details on the annotation pipeline and quality labels, please see Appendix V.

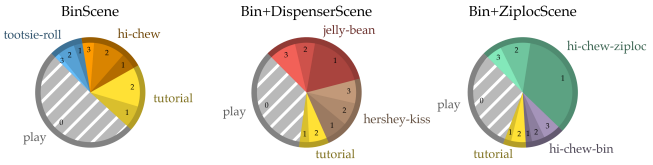


Fig. 4: **Dataset composition by number of time steps for each of our three scenes**. Different hues indicate different tasks. Tasks receive quality scores from 1 to 3 (higher is better) which are also indicated by brighter shades. Tutorial data receives a score of 1 or 2. Play data always receives a score of 0.

Metrics. We analyze the crowdsourced data using several metrics:

- *Quantity.* Our primary metric measures the number of timesteps a user spends performing a task.
- *Quality*. We utilize our data quality annotations, and additionally explore other data quality measures in Appendix III.
- *Usefulness for Policy Learning.* We study the utility of the crowdsourced data for policy learning via co-training and fine-tuning with expert demonstrations.
- *Self-reported Likert Ratings.* We survey users for self-ratings of intuitiveness, enjoyment, and how well the robot completed the task in the way they desired, and report results in Appendix III.

VI. Results

In this section, we analyze the composition of the dataset, the effects of different incentive mechanisms, and the usefulness of the data for policy learning.

A. Usage Overview and Dataset Composition

We observe significant engagement with RoboCrowd over the two-week collection period: there were $N = 231$ unique users in total. On most days, more than two-thirds of these were new users that had not used the system on prior days. There were a total of 814 interaction episodes distributed throughout the period.

Our dataset is composed of 3 scenes (Fig. 3). We collect 129 interaction episodes in `BinScene` (Day 1), 381 in `Bin+DispenserScene` (Days 2-5), and 307 in `Bin+ZiplocScene` (Days 6-11). In aggregate, users spent 54.2% of interaction time performing the preset tasks in the scene, 9.6% on the interactive tutorial, and 36.1% on free-play. While we focus our learning experiments on task-relevant data in Section VI-C, this play data could be fruitful for training multitask policies in the future. In Fig. 4, we show the distribution of tasks and qualities over timesteps for each scene. Qualities are determined on a scale from 1–3 for task-relevant data and a scale of 1–2 for tutorial data based on the smoothness of the user's motion and whether there is retrying behavior or extraneous movements. We detail the quality annotation rules in Appendix V, and illustrate sample trajectories in each scene in Appendix I.

B. Effects of Incentives on Data Quantity and Quantity

Material Rewards. While `BinScene` contains two bin-picking tasks with nearly identical difficulty, users in aggregate spend 2× as many timesteps performing `hi-chew` compared to `tootsie-roll`. This suggests that users devote more interaction time to tasks where the direct material incentive is more preferred. We also see that users spend a significant amount of time (50.7%) in free-play with the system in `BinScene`, engaging in behaviors such as trying out more challenging tasks (e.g., attempting to

Task	Scene	# Exp.	Expert	Co-train	Fine-tune
hi-chew	B	30	37.5%	27.5%	**42.5%**
tootsie-roll	B	30	**42.5%**	25%	40%
hershey-kiss	B+D	60	20%	32.5%	**35%**
hi-chew-bin	B+D	80	20%	12.5%	**40%**
jelly-bean	B+Z	100	**48.9 ± 18.6**	8.9 ± 10.1	19.7 ± 29.7
hi-chew-ziploc	B+Z	100	5.4 ± 12.2	17.1 ± 15.8	**22.1 ± 14.3**

TABLE I: **Policy Performance.** Performance of policies trained on expert demonstrations (# Exp.), co-trained on crowd data, and pre-trained on expert+crowd data then fine-tuned on expert data. We conduct 40 trials for each cell. For the long-horizon tasks (jelly-bean, hi-chew-ziploc), we provide a normalized return (out of 100) rather than success rate (see Appendix IV for details).

unwrap the candies; see Appendix II). Thus, while material incentives can influence user demonstrations (e.g., higher material incentives can lead to more data), drivers of intrinsic motivation such as the difficulty of the task also play a role, as we discuss next.

Intrinsic Motivation. Interestingly, in Bin+DispenserScene, which contains a harder bin-picking task than in Scene A (hershey-kiss) and a challenging long-horizon candy dispensing task (jelly-bean), users spend only 35.3% of the time in free-play. Additionally, despite the fact that users do not generally prefer Jelly Beans over Hershey Kisses as a material reward, they still spend more (1.5×) time performing the jelly-bean task. This suggests that intrinsic interest can influence users to allocate more time doing harder task compared to easier ones, or engaging in free-play. To probe whether this intrinsic motivation effect is present even when controlling for the material reward, we consider Bin+ZiplocScene. Here, the incentive is contained within a closed Ziploc bag which must be opened. The same incentive is available in the bin to be picked. Users spend 4.18× as many timesteps on hi-chew-ziploc compared to hi-chew-bin, again suggesting that intrinsic motivation influences which tasks users perform in the scene.

Social Comparison. To examine how different people respond differently to explicit comparison mechanisms in the system, we record which users visit the Leaderboard Page, and conduct a Mann-Whitney U-test to compare the quantity and quality of demonstrations provided by Leaderboard visitors compared to other users. Fig. 5 illustrates the distribution of quality (number of interactions) and quality (mean quality score) conditioned on Leaderboard visitation. We find that that visitors of the Leaderboard provide significantly more demonstrations ($p < 0.001$) that are higher quality on average ($p < 0.05$).

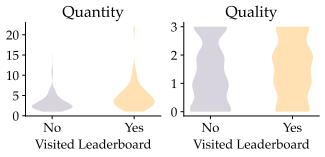


Fig. 5: **Quantity and quality by leaderboard use**. Violin plot showing the distribution of quantity and quality of demonstrations for users who did and did not visit the leaderboard.

C. Policy Learning with the Crowdsourced Data

In this section, we study how useful the crowdsourced data is for downstream policy learning. To complement the crowdsourced data, we collect a set of high-quality expert demonstrations for each task: 30 demonstrations for each of hi-chew and tootsie-roll, 60 for hershey-kiss, 80 for hi-chew-bin, and 100 for each of jelly-bean and hi-chew-ziploc.

In Table I, we compare different methods of mixing crowdsourced data and expert data on our six tasks. All policies use ACT [1] with default hyperparameters. Training exclusively with the expert data on each task constitutes the *Expert* setting. *Co-train* refers to naïvely mixing data from a crowdsourced task (i.e., task-relevant data of any quality) with the expert data. We also compare to *Fine-tune*, which trains in two stages: first co-training on the crowd data and expert data and then fine-tuning on expert data only; for fair comparison, note that *Fine-tune* is trained for fewer total steps (150K) than both *Expert* and *Co-train* (200K). In most cases, the crowdsourced data provides performance improvements, especially for more complex tasks, but the specific results vary by task. For example, crowdsourced data for the bin-picking tasks can involve low-quality behaviors (i.e., regrasping behavior or grasping multiple items at a time), which may cause the *Co-train* to perform worse than *Expert*, but still provide a useful initialization for *Fine-tune*. We provide additional qualitative analysis of the trained policies in Appendix B.

We also demonstrate that the crowdsourced data can benefit downstream tasks. In Table II, we train an expert-only ACT policy (50 demos) to convergence on a new task, tool-ziploc, which requires unzipping a Ziploc containing tools. The unzipping skill is shared with hi-chew-ziploc. We compare this expert-only policy to a policy trained with two stages (pre-training on crowdsourced hi-chew-ziploc data and then fine-tuning on the expert tool-ziploc data). This outperforms the expert-only policy by 20%, suggesting that crowdsourced data can be beneficial in downstream tasks with shared manipulation skills.

Subtask	Expert	Fine-tune
Grasp Corner	85%	**95%**
Pinch Tab	70%	**85%**
Slide Open	60%	**80%**

TABLE II: Staged success rate for a policy pre-trained on hi-chew-ziploc crowd data and fine-tuned on 50 expert demos of tool-ziploc compared to expert-only tool-ziploc policy.

VII. Discussion and Limitations

In this work, we propose a new paradigm for robot data collection via crowdsourcing and incentive design. We focus on three incentive types—material rewards, intrinsic motivation, and social comparison—but there are further avenues to explore within these categories as well (e.g., how physical rewards differ from monetary incentives). Crafting data collection schemes where people are motivated by external rewards, fun, interest, or competition is a general principle, and a rich area for future work would be to scale up our findings on incentive design in robot data collection to new tasks. For example, appealing to extrinsic motivation and social comparison could help craft a data collection scheme for a task such as packing groceries—where users are motivated by spaced rewards (getting to keep every N bags) or social comparison (getting points for more efficient packing). A variety of other incentive types (e.g., task novelty, collective effort, robot's ability to learn from the data, etc.) could be applied to new settings as well. While crowdsourcing has the benefit of reducing data collection effort of individual researchers, it also presents challenges of data quality and heterogeneity. We hope that our dataset—collected from over 200 users with manual fine-grained quality annotations—can be helpful to future works seeking to understand the style and diversity of different human operators, and what the most effective ways are to leverage crowdsourced data during downstream policy learning.

References

[1] T. Zhao, V. Kumar, S. Levine, and C. Finn, "Learning Fine-Grained Bimanual Manipulation with Low-Cost Hardware," in *Proceedings of Robotics: Science and Systems (RSS)*, 2023.

[2] A. Brohan, N. Brown, J. Carbajal, *et al.*, "RT-1: Robotics Transformer for Real-World Control at Scale," in *arXiv*, 2022.

[3] A. Brohan, N. Brown, J. Carbajal, *et al.*, "RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control," in *arXiv*, 2023.

[4] Octo Model Team, D. Ghosh, H. Walke, *et al.*, "Octo: An Open-Source Generalist Robot Policy," in *Proceedings of Robotics: Science and Systems (RSS)*, 2024.

[5] M. Kim, K. Pertsch, S. Karamcheti, *et al.*, "Openvla: An open-source vision-language-action model," *arXiv*, 2024.

[6] O. X.-E. Collaboration, A. O'Neill, A. Rehman, *et al.*, *Open X-Embodiment: Robotic Learning Datasets and RT-X Models*, 2023.

[7] Y. J. Ma, S. Sodhani, D. Jayaraman, O. Bastani, V. Kumar, and A. Zhang, "VIP: Towards Universal Visual Reward and Representation via Value-Implicit Pre-Training," *International Conference on Learning Representations (ICLR)*, 2023.

[8] S. Nair, A. Rajeswaran, V. Kumar, C. Finn, and A. Gupta, "R3M: A Universal Visual Representation for Robot Manipulation," in *Conference on Robot Learning (CoRL)*, 2022.

[9] A. Khazatsky, K. Pertsch, S. Nair, *et al.*, "DROID: A Large-Scale In-The-Wild Robot Manipulation Dataset," *arXiv*, 2024.

[10] S. Dasari, F. Ebert, S. Tian, *et al.*, "RoboNet: Large-Scale Multi-Robot Learning," in *Conference on Robot Learning (CoRL)*, 2019.

[11] S. Nair, A. Rajeswaran, V. Kumar, C. Finn, and A. Gupta, "R3M: A universal visual representation for robot manipulation," in *Conference on Robot Learning (CoRL)*, 2022.

[12] I. Radosavovic, T. Xiao, S. James, P. Abbeel, J. Malik, and T. Darrell, "Real-world robot learning with masked visual pre-training," in *Conference on Robot Learning (CoRL)*, 2023.

[13] S. Karamcheti, S. Nair, A. S. Chen, *et al.*, "Language-Driven Representation Learning for Robotics," in *Proceedings of Robotics: Science and Systems (RSS)*, 2023.

[14] S. Ross, G. J. Gordon, and J. A. Bagnell, "A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning," in *International Conference on Artificial Intelligence and Statistics*, 2010.

[15] M. Kelly, C. Sidrane, K. Driggs-Campbell, and M. J. Kochenderfer, "HG-DAgger: Interactive Imitation Learning with Human Experts," *International Conference on Robotics and Automation (ICRA)*, 2018.

[16] K. Gandhi, S. Karamcheti, M. Liao, and D. Sadigh, "Eliciting Compatible Demonstrations for Multi-Human Imitation Learning," in *Proceedings of the 6th Conference on Robot Learning (CoRL)*, 2022.

[17] R. Snow, B. O'Connor, D. Jurafsky, and A. Y. Ng, "Cheap and Fast - But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks," in *Conference on Empirical Methods in Natural Language Processing*, 2008.

[18] A. Sorokin and D. A. Forsyth, "Utility data annotation with Amazon Mechanical Turk," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[19] L. von Ahn and L. Dabbish, "Labeling images with a computer game," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2004.

[20] M. S. Bernstein, D. S. Tan, G. Smith, M. Czerwinski, and E. Horvitz, "Collabio: a game for annotating people within social networks," in *ACM Symposium on User Interface Software and Technology (UIST)*, 2009.

[21] J. Park, R. Krishna, P. Khadpe, L. Fei-Fei, and M. S. Bernstein, "AI-Based Request Augmentation to Increase Crowdsourcing Participation," in *Proceedings of the Seventh AAAI Conference on Human Computation and Crowdsourcing, HCOMP*, 2019.

[22] R. Krishna, Y. Zhu, O. Groth, *et al.*, "Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations," *International Journal of Computer Vision*, 2016.

[23] N. Zhou, Z. D. Siegel, S. Zarecor, *et al.*, "Crowdsourcing image analysis for plant phenomics to generate ground truth data for machine learning," *PLOS Computational Biology*, Jul. 2018.

[24] S. Ørting, A. Doyle, A. van Hilten, *et al.*, "A Survey of Crowdsourcing in Medical Image Analysis," *Human Computation*, 2019.

[25] T. W. Cenggoro, F. Tanzil, A. H. Aslamiah, E. K. Karuppiah, and B. Pardamean, "Crowdsourcing annotation system of object counting dataset for deep learning algorithm," *IOP Conference Series: Earth and Environmental Science*, 2018.

[26] M. van Vliet, E. C. Groen, F. Dalpiaz, and S. Brinkkemper, "Identifying and Classifying User Requirements in Online Feedback via Crowdsourcing," in *Requirements Engineering: Foundation for Software Quality*, 2020.

[27] B. Shmueli, J. Fell, S. Ray, and L.-W. Ku, "Beyond Fair Pay: Ethical Implications of NLP Crowdsourcing," in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics, 2021.

[28] N. Nangia, S. Sugawara, H. Trivedi, A. Warstadt, C. Vania, and S. R. Bowman, "What Ingredients Make for an Effective Crowdsourcing Protocol for Difficult NLU Data Collection Tasks?" *arXiv*, 2021.

[29] S. Mishra, D. Khashabi, C. Baral, and H. Hajishirzi, "Cross-Task Generalization via Natural Language Crowdsourcing Instructions," *arXiv*, 2022.

[30] S. Lim, A. Jatowt, M. Färber, and M. Yoshikawa, "Annotating and Analyzing Biased Sentences in News Articles using Crowdsourcing," in *Proceedings of the Twelfth*

Language Resources and Evaluation Conference, May 2020.

[31] L. B. Chilton, G. Little, D. Edge, D. S. Weld, and J. A. Landay, "Cascade: crowdsourcing taxonomy creation," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2013.

[32] K. Crowston, "Amazon Mechanical Turk: A Research Tool for Organizations and Information Systems Scholars," in *Shaping the Future of ICT Research. Methods and Approaches*, 2012.

[33] S. Palan and C. Schitter, "Prolific. ac—A subject pool for online experiments," *Journal of Behavioral and Experimental Finance*, 2018.

[34] L. Von Ahn, B. Maurer, C. McMillen, D. Abraham, and M. Blum, "recaptcha: Human-based character recognition via web security measures," *Science*, 2008.

[35] E. Law and L. Von Ahn, *Human computation*. Morgan & Claypool Publishers, 2011.

[36] H. Garcia-Molina, M. Joglekar, A. Marcus, A. Parameswaran, and V. Verroios, "Challenges in Data Crowdsourcing," *IEEE Transactions on Knowledge and Data Engineering*, 2016.

[37] J. Huynh, J. Bigham, and M. Eskenazi, "A Survey of NLP-Related Crowdsourcing HITs: what works and what does not," *arXiv*, 2021.

[38] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From Game Design Elements to Gamefulness: Defining "Gamification"," in *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, 2011.

[39] A. Mandlekar, Y. Zhu, A. Garg, *et al.*, "RoboTurk: A Crowdsourcing Platform for Robotic Skill Learning through Imitation," in *Conference on Robot Learning*, 2018.

[40] A. Mandlekar, J. Booher, M. Spero, *et al.*, "Scaling robot supervision to hundreds of hours with roboturk: Robotic manipulation dataset through human reasoning and dexterity," in *International Conference on Intelligent Robots and Systems (IROS)*, 2019.

[41] M. Balsells, M. M. L. Torné, Z. Wang, S. Desai, P. Agrawal, and A. Gupta, "Autonomous Robotic Reinforcement Learning with Asynchronous Human Feedback," *arXiv*, 2023.

[42] M. Torne, M. Balsells, Z. Wang, *et al.*, "Breadcrumbs to the Goal: Goal-Conditioned Exploration from Human-in-the-Loop Feedback," *arXiv*, 2023.

[43] M. J.-Y. Chung, M. Forbes, M. Cakmak, and R. P. Rao, "Accelerating imitation learning through crowdsourcing," in *International Conference on Robotics and Automation (ICRA)*, 2014.

[44] S. Van Waveren, E. J. Carter, O. Örnberg, and I. Leite, "Exploring non-expert robot programming through crowdsourcing," *Frontiers in Robotics and AI*, 2021.

[45] S. Chernova, N. DePalma, E. Morant, and C. Breazeal, "Crowdsourcing human-robot interaction: Application from virtual to physical worlds," in *2011 RO-MAN*, 2011.

[46] S. Song, A. Zeng, J. Lee, and T. Funkhouser, "Grasping in the wild: Learning 6dof closed-loop grasping from low-cost demonstrations," *IEEE Robotics and Automation Letters (RA-L)*, 2020.

[47] C. Wang, H. Shi, W. Wang, R. Zhang, L. Fei-Fei, and C. K. Liu, "DexCap: Scalable and Portable Mocap Data Collection System for Dexterous Manipulation," *arXiv*, 2024.

[48] C. Chi, Z. Xu, C. Pan, *et al.*, "Universal Manipulation Interface: In-The-Wild Robot Teaching Without In-The-Wild Robots," *arXiv*, 2024.

[49] J. Wang, C.-C. Chang, J. Duan, D. Fox, and R. Krishna, "Eve: Enabling anyone to train robot using augmented reality," *ACM Symposium on User Interface Software and Technology (UIST)*, 2024.

[50] L. J. Ratliff, R. Dong, S. Sekar, and T. Fiez, "A perspective on incentive design: Challenges and opportunities," *Annual Review of Control, Robotics, and Autonomous Systems*, 2019.

[51] A. W. Kruglanski and O. Mayseless, "Classic and current social comparison research: Expanding the perspective.," *Psychological bulletin*, 1990.

[52] E. Todorov, T. Erez, and Y. Tassa, "Mujoco: A physics engine for model-based control," in *International Conference on Intelligent Robots and Systems (IROS)*, 2012.

[53] B. Medler and B. Magerko, "Analytics of play: Using information visualization and gameplay practices for visualizing video game data," *Parsons Journal for Information Mapping*, 2011.

[54] Z. Fu, T. Zhao, and C. Finn, "Mobile ALOHA: Learning Bimanual Mobile Manipulation with Low-Cost Whole-Body Teleoperation," *arXiv*, 2024.

APPENDIX

OVERVIEW

In the appendices below, we provide additional details on the implementation of RoboCrowd, our experiments, and our crowdsourced dataset. We provide a brief overview of each appendix below. For videos, please see our website: https://robocrowd.github.io

Appendix I – *Task Details*

We give descriptions of each of our 6 tasks, as well as renderings and images depicting sample expert demonstrations for each task.

Appendix II – *Dataset Examples*

We provide sample trajectories from our collected dataset including their task and quality annotations, to qualitatively illustrate the diversity of the behaviors in the dataset.

Appendix III – *Additional Dataset Analysis*

We provide further data analysis, including an offline user study to justify our scene choices, additional data quality analysis, and results on users' self-reported Likert ratings of their interactions with the system.

Appendix IV – *Additional Details on Policy Learning Experiments*

We provide additional details on the training and evaluation procedures for our policy learning experiments, as well as further qualitative analysis of the results.

Appendix V – *Additional Details on Software Implementation and Data Annotation*

We provide further details on the graphical user interface, interactive tutorial, software implementation, and data annotation pipeline.

Appendix VI – *Additional Details on Pilot Studies and System Development*

We provide more details on how we designed and refined the system through pilot studies.

Appendix VII – *Overview of Action Chunking with Transfomers (ACT) [1]*

We provide additional background on the Action Chunking with Transfomers (ACT) algorithm.

APPENDIX I
TASK DETAILS

In Tables III to VIII below, we provide a verbal description of the behavior that the expert demonstrations perform for each task. We additionally include a virtual rendering of different segments of a sample demonstration (where the gripper is rendered with increasing opacity for later timesteps). Additionally, we show a timelapse of the overhead camera image observation for the same sample expert demonstration.

APPENDIX II
DATASET EXAMPLES

In Figs. 6 to 8, we give 3 qualitative examples of interaction episodes in our crowdsourced dataset. We illustrate a timelapse of each episode with the overhead camera observation. We also include the task and quality annotations at each timestep, with a verbal description of the episode in the caption.

APPENDIX III
ADDITIONAL DATASET ANALYSIS

In this section, we provide additional data analysis. In Appendix A, we describe an offline study over user preferences for different candies, informing our different scene setups. In Appendix B and Appendix B.1, we examine additional metrics (i.e., tutorial quality and Likert ratings) that correlate with quality of user interaction episodes, and in Appendix C, we provide additional statistics on usage and retention.

A. Justification for Scene Choices

To justify our scene setup and task pairings, we perform an offline survey on user preferences for various candies. On a sample of $N = 16$ users, we find that 81% prefer a Hi-Chew to a Tootsie Roll. Thus, `BinScene` (which includes the `hi-chew` and `tootsie-roll` tasks) allows us investigate whether this preference for material reward shapes task choice when teleoperating demonstrations, when the task is otherwise equivalent besides the material reward. Users exhibit a more mild preference for a Hershey Kiss compared to a small handful of Jelly Beans (with 62% of respondents preferring the Hershey Kiss). `Bin+ZiplocScene` (which includes the `hi-chew-bin` and `hi-chew-ziploc` tasks) allows us to investigate how intrinsic motivation and task difficulty affects user behavior when teleoperating in the case that the material reward (a Hi-Chew) is held constant between the the simpler task and the more challenging task. `Bin+DispenserScene` allows us to investigate this question when the material rewards are different, and users do not exhibit an overall preference for the reward from the harder task (and even mildly prefer the reward from the easier task).

B. Additional Metrics on Demonstration Quality

Our crowdsourced dataset contains rich interaction data per user ID—during and after the interactive tutorial period. This dataset can help to yield insights about which users give higher quality trajectories, and what factors can help predict this quality. As an example, we examine how the quality of interactions *after* the tutorial (i.e., when the user selects tasks in the scene to perform) correlates with quality *during* the tutorial period (i.e., when the user is instructed to complete simple onboarding tasks). Specifically, we examine the distribution of mean quality during task interactions versus minimum quality during the tutorial period; the user's tutorial period is classified as 0 if there is any off-task behavior, 1 if the tutorial is performed but with retrying, and 2 if the tutorial is performed smoothly. We observe a loose positive correlation between higher minimum tutorial quality and mean task quality; and notably, users who produce consistently high quality task demonstrations (quality 3) are more present in the group with high quality tutorials. The tutorial period can therefore be a first-cut proxy at filtering demonstrators by quality.

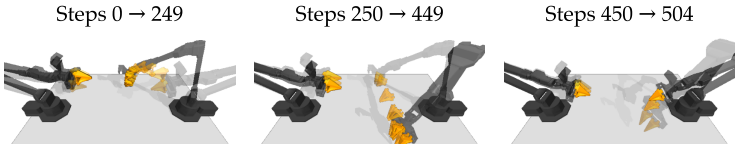
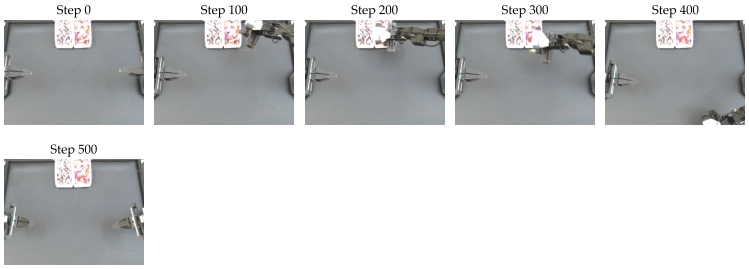
Task Name	Pick up a Hi-Chew (`hi-chew`)
Task Description	Move the right arm towards the candy bin. Grasp one Hi-Chew. Drop it in the End Zone. Finally, return to the home position.
Expert Trajectory Rendering	
Expert Trajectory Timelapse	

TABLE III: Description of the `hi-chew` task, as well as a rendering and timelapse of a sample expert trajectory.

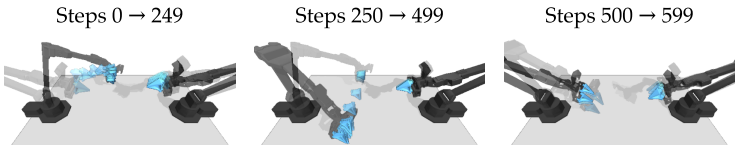
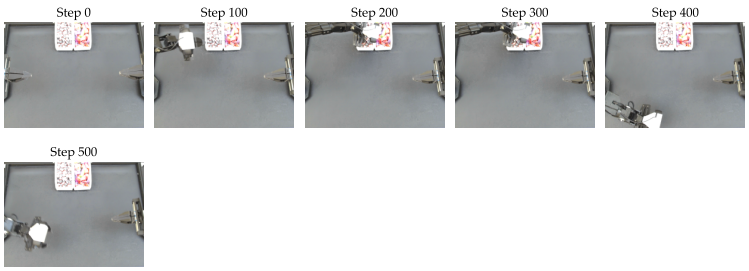
Task Name	Pick up a Tootsie Roll (`tootsie-roll`)
Task Description	Move the left arm towards the candy bin. Grasp one Tootsie Roll. Drop it in the End Zone. Finally, return to the home position.
Expert Trajectory Rendering	
Expert Trajectory Timelapse	

TABLE IV: Description of the `tootsie-roll` task, as well as a rendering and timelapse of a sample expert trajectory.

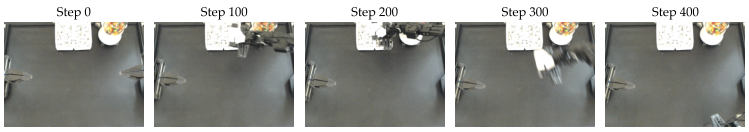
Task Name	Pick up a Hershey Kiss (`hershey-kiss`)
Task Description	Move the right arm **or** the left arm towards the candy bin. Grasp one Hershey Kiss. Drop it in the End Zone. Finally, return to the home position.
Expert Trajectory Rendering	
Expert Trajectory Timelapse	

TABLE V: Description of the `hershey-kiss` task, as well as a rendering and timelapse of a sample expert trajectory.

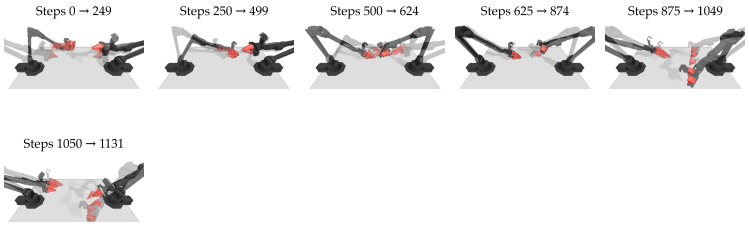
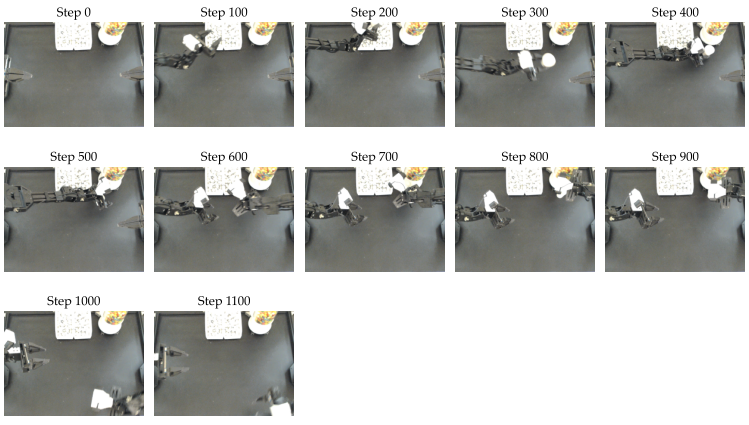
Task Name	Eject a Jelly Bean from the Candy Dispenser (`jelly-bean`)
Task Description	Use the left arm to pull a cup from the cup dispenser. Bring the cup near the lever of the candy dispenser. Use the right arm to align the cup under the lever, then press the lever. Then, use the right arm to pick up the cup and bring it to the End Zone. Finally, return to the home position.
Expert Trajectory Rendering	
Expert Trajectory Timelapse	

TABLE VI: Description of the `jelly-bean` task, as well as a rendering and timelapse of a sample expert trajectory.

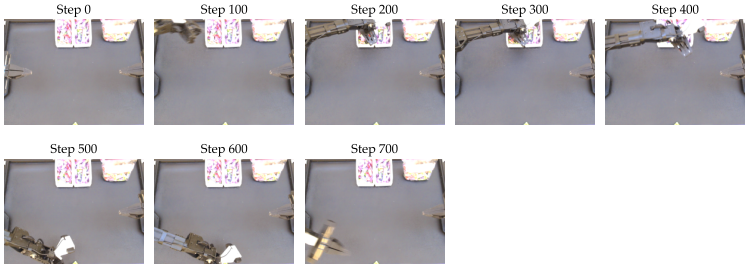
Task Name	Pick up a Hi-Chew from the Bin (`hi-chew-bin`)
Task Description	Move the right arm **or** the left arm towards the candy bin. Grasp one Hi-Chew. Drop it in the End Zone. Finally, return to the home position.
Expert Trajectory Rendering	
Expert Trajectory Timelapse	

TABLE VII: Description of the `hi-chew-bin` task, as well as a rendering and timelapse of a sample expert trajectory.

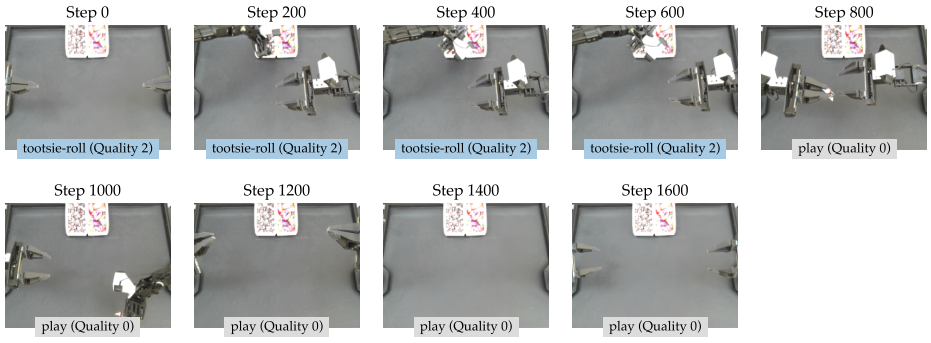


Fig. 6: In this trajectory, the user begins by performing the `tootsie-roll` task with moderate quality—i.e., there are about 3 attempts to grasp the candy, and there is some extraneous movement in the right arm, but the user is otherwise successful at grasping the candy. Before bringing the candy all the way to the End Zone, the user attempts to unwrap the candy. They then hand it over to the other arm, place it in the End Zone, and then move the arms upward. The first half of the episode is marked as `tootsie-roll` (Quality 2) and the latter half of the episode is marked as `play` (Quality 0).

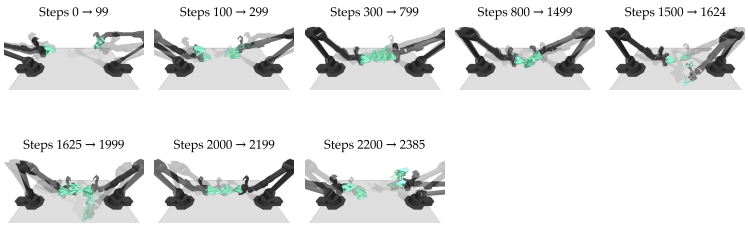
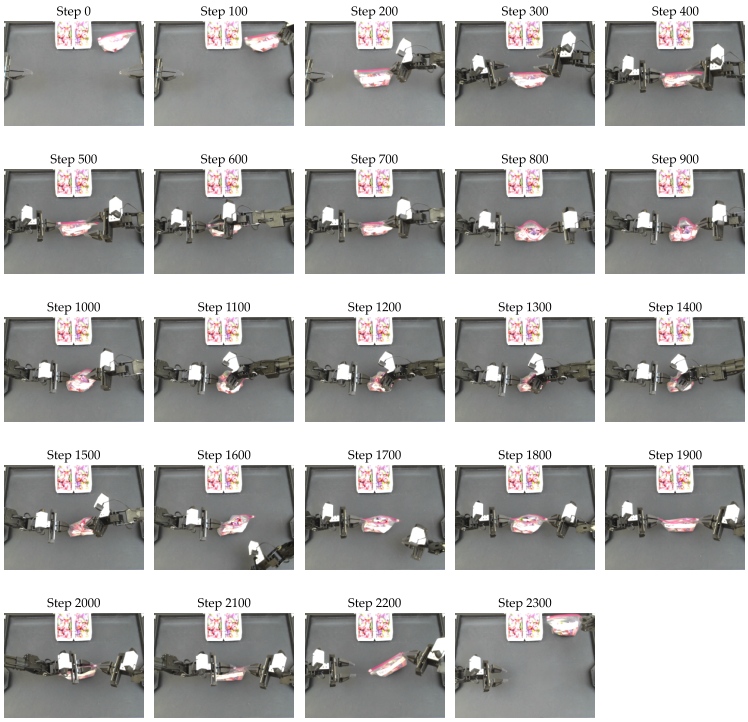
Task Name	Open the Ziploc, Pick up a Hi-Chew, then Close the Ziploc (`hi-chew-ziploc`)
Task Description	Use the right arm to bring the Ziploc bag to the center of the table. Then, use the left arm to hold the Ziploc while pulling the Ziploc tab with the right arm to open the bag. Then, spread the Ziploc open and pick out a Hi-Chew with the right arm, and bring it to the End Zone. Then, use the right arm to hold the Ziploc while pulling the Ziploc tab closed with the left arm. Finally, use the right arm to place the Ziploc back in the corner of the table, and return the arms to the home position.
Expert Trajectory Rendering	
Expert Trajectory Timelapse	

TABLE VIII: Description of the `hi-chew-ziploc` task, as well as a rendering and timelapse of a sample expert trajectory.

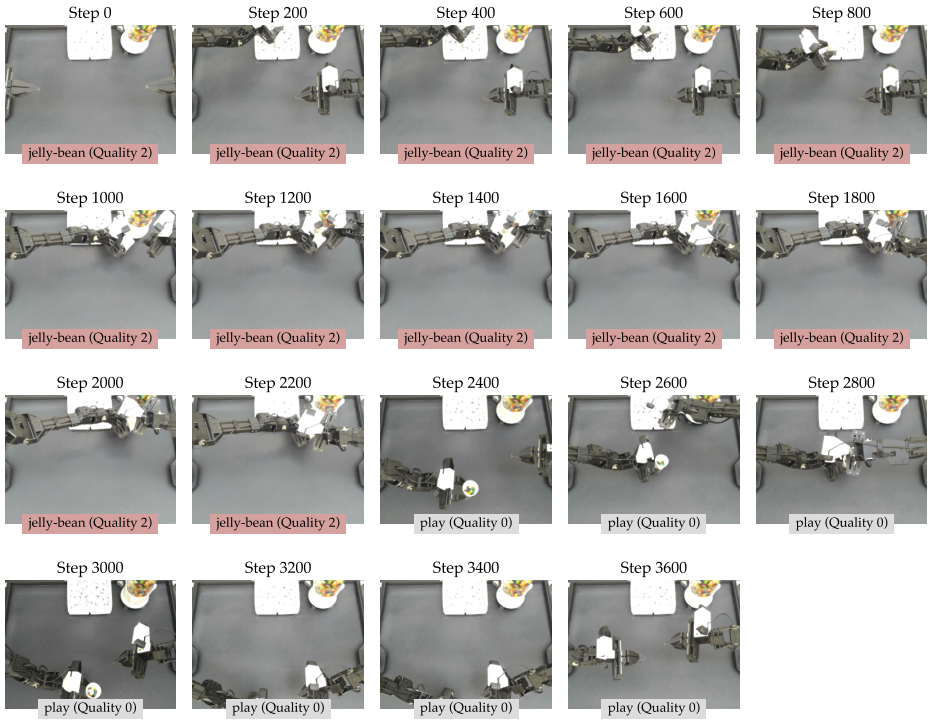


Fig. 7: In this trajectory, the user grasps a cup from the cup dispenser and places it under the lever of the candy machine. They are successful in collecting jelly beans in the cup, though the trajectory includes retrying behavior and is not as smooth as an expert trajectory. The user brings the cup halfway to the End Zone, and then begins behaviors that are not part of the task—i.e., placing a Hershey Kiss in the cup before bringing it to the End Zone. The first part of the episode is marked as `jelly-bean` (Quality 2) and the latter part is marked as `play` (Quality 0).

Fig. 8: In this trajectory, the user correctly moves the Ziploc from the corner of the table to the center of the table, and grasps a Hi-Chew from inside the Ziploc which they bring to the End Zone. They are unsuccessful in closing the Ziploc before episode termination. The user is task-directed for the whole episode, however takes longer than better quality trajectories for this task and performs retrying behavior at each subtask. The whole trajectory is marked as `hi-chew-ziploc` (Quality 1).

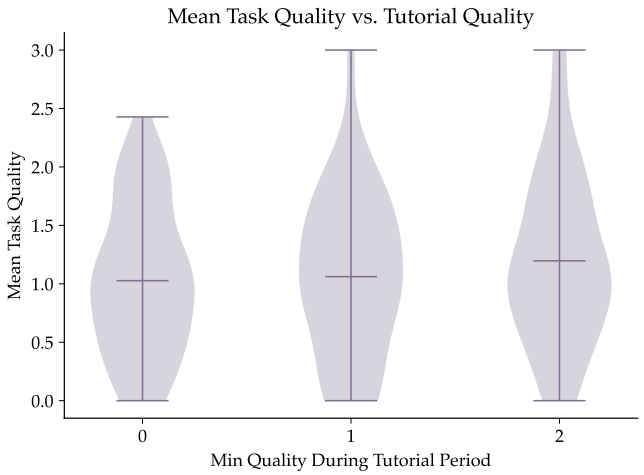


Fig. 9: Distribution of Mean Task Quality versus Minimum Quality during the Tutorial Period.

Learning Rate	1e-5
Batch Size	8
# Encoder Layers	4
# Decoder Layers	7
Feedforward Dimension	3200
Hidden Dimension	512
# Heads	8
Chunk Size	100
KL-weight (β)	10
Dropout	0.1
Backbone	ResNet-18
Image Augmentations	RandomCrop, RandomResize, RandomRotation, ColorJitter

TABLE IX: Hyperparameters for ACT, shared for all experiments.

1) Self-Reported Likert Metrics: After every interaction episode, we prompt the user to answer whether they agree with 3 statements, on a 5-point scale (1 - Strongly Disagree; 2 - Disagree; 3 - Neutral; 4 - Agree; 5 - Strongly Agree).

- `Intuitive`: Controlling the robot was intuitive.
- `Interesting`: Controlling the robot was fun and interesting.
- `Wanted`: The robot accomplished the task in the way that I wanted.

Fig. 10 summarizes the responses to these questions, aggregated by users' minimum ratings to each statement over their interaction episodes. The majority of users agree with all three statements, and most often have the strongest ratings for `Interesting` compared to `Intuitive` and `Wanted`. We find also that there are loose correlations between the manually annotated quality scores for users' interaction episodes and users' self-reported ratings for each of these metrics. Specifically, users who self-report low ratings on each of the three metrics have lower mean quality scores. However, users who self-report high ratings have quality scores that span low to high.

C. Usage and Retention

We illustrate the usage of the RoboCrowd in Fig. 11. We observe significant engagement with RoboCrowd over the two-week collection period: there were $N = 231$ unique users in total. On most days, more than two-thirds of these were new users that had not used the system on prior days. There were a total of 814 interaction episodes distributed throughout the period. The most common time at which users interacted with the system was about 1pm, corresponding to the most trafficked time in the café (lunchtime). We collect 129 interaction episodes in `BinScene` (Day 1), 381 in `Bin+DispenserScene` (Days 2-5), and 307 in `Bin+ZiplocScene` (Days 6-11).

Appendix IV
Additional Details on Policy Learning Experiments

In this section, we give additional details on our policy learning experiments. Appendix A provides training details and hyperparameters, Appendix B provides details on our evaluation procedure, and Appendix C provides additional qualitative discussion of our learned policies.

A. Training Details

For the *Expert* and *Co-train* experiments, we train policies for 200K steps for all tasks. For the *Fine-tune* experiments, we fine-tune the co-trained model (partially trained for 100K steps) for an additional 50K steps on expert data only. We use the implementation of ACT [1] from [54], including the default hyperparameters from [1], as shown in Table IX.

B. Evaluation Details

We perform policy evaluations for 40 trials each, early stopping when policies exhibit excessively jittery or unsafe behavior. While the RoboCrowd training dataset was collected in a café where lighting varies throughout the day, during evaluation, we move the setup to a location with a visually similar background but consistent lighting for controlled evaluations.

For the bin-picking tasks, we define success as the robot arm picking exactly one of the desired candy and bringing it to the End Zone. For our challenging, long-horizon tasks (`jelly-bean` and `hi-chew-ziploc`), success is 0% for all policies, so we instead compare policies via normalized return to measure partial proficiency at tasks. We describe the process for computing normalized return below.

Each of the following subtasks in `jelly-bean` corresponds to 1 point in the episode return: Retrieves Cup from Dispenser; Places Cup Down; Aligns Cup Under Lever; Presses Lever; Collects Jelly Beans in Cup; Picks up Cup; Brings Cup to End Zone. Each of the following subtasks in `hi-chew-ziploc` corresponds to 1 point in the episode return: Picks up Bag; Places Bag in Center of Table; Slides Open; Picks Hi-Chew; Brings Hi-Chew to End Zone; Closes Bag; Places Bag in Corner of Table. For these tasks, we report normalized return—the average return over evaluation trials divided by the maximum return (achieved by all expert demonstrations).

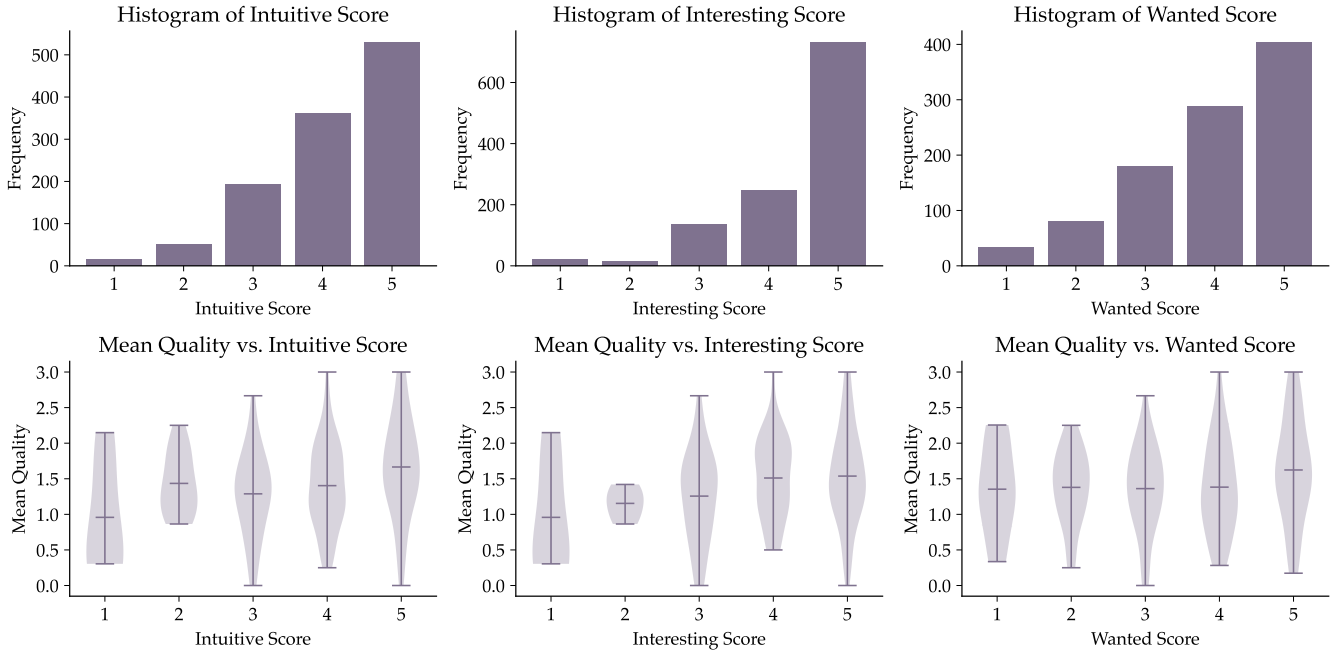


Fig. 10: (*Top*) Histogram of Likert Ratings (aggregated by the user's minimum response over their interaction episodes) for the `Intuitive`, `Interesting`, and `Wanted` questions. (*Bottom*) Distribution of mean quality of interaction episodes for different Likert Ratings for `Intuitive`, `Interesting`, and `Wanted`.

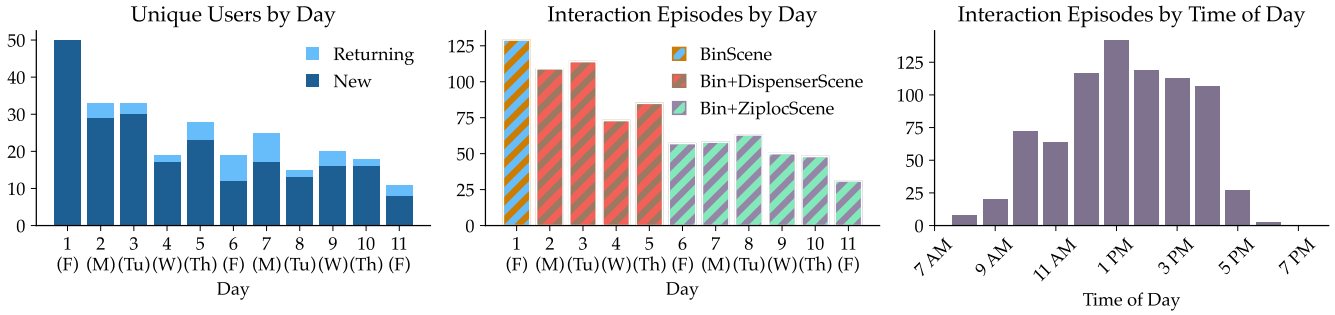


Fig. 11: Statistics on usage over a two-week period: number of users per day (left), number of interaction episodes per day (middle), and distribution of interaction episodes by time of day (right).

C. *Qualitative Analysis of Learned Policies*

We find that in most cases, *Co-train* and/or *Fine-tune* improve upon *Expert*. However, the specific effects vary by task. For example, we find that for the `hi-chew` task, the co-trained policy performs worse than the expert policy, but the fine-tuned policy performs better; whereas with the `hershey-kiss` task, both the co-trained policy and fine-tuned policy perform better. We hypothesize that the crowdsourced data is more useful for `hershey-kiss` because (a) `hershey-kiss` is a more complex task (in that it is more multimodal, i.e., either arm can be used to pick up a Hershey Kiss, and the grasping required needs to be more precise to not crush the Hershey Kiss) and (b) a greater proportion of the `hershey-kiss` data is of higher quality. We notice that the crowdsourced data for `jelly-bean` is especially diverse, and naïvely co-training or fine-tuning underperforms using the expert data only.

Qualitatively, we observe in several cases that the co-trained and fine-tune policies exhibit meaningful but suboptimal behaviors from the crowdsourced data (e.g., picking up multiple objects from the bin instead of one). On the other hand, there are also helpful behaviors from the crowdsourced data (*not* represented in the expert data) that benefit trained policies—e.g., regrasping behavior.

Overall, the RoboCrowd dataset is very diverse, and contains both task-relevant behaviors (of various levels of quality) and free-play behavior. Future work on more sophisticated policy learning methods that leverage these diverse characteristics can help to get the maximum utility out of crowdsourced demonstration data.

APPENDIX V
ADDITIONAL DETAILS ON SOFTWARE IMPLEMENTATION AND DATA ANNOTATION

In this section, we provide additional details on our software interface and implementation, as well as our data annotation pipeline. Appendix A provides an overview of the application flow and interface, Appendix B details the interactive tutorial procedure, Appendix C provides implementation details, and Appendix D details the data annotation pipeline.

A. Application Flow and User Interface

Fig. 12 gives an overview of the flow through the tablet application, and Table X provides screenshots of the major pages referenced in the flowchart. We additionally highlight the Interactive Tutorial in Fig. 13 and the visual warning for collision detection in Fig. 14. We now briefly describe the application flow. To begin a new session, the user taps their ID card on the card reader, which advances the tablet application to a screen where the user can enter a nickname (if they are a new user). They are then directed to the Main Page, where they complete a consent form and the interactive tutorial. From the Main Page, users can also press a "Start Playing" button which directs them to the Task Page, where they can see videos of tasks available in the scene, and can tap on a task to see more details and begin demonstrating the task. For safety, the user receives an audial and visual warning (Fig. 14) if the arms are near-collision. When users are done with the task (i.e., they click a Stop button on the Task Detail Page or they rest the grippers on the mechanical stop), they are asked to mark their demonstration as a success or failure, and fill out a brief survey. The success/failure markings are used as the basis for the points which are added to the user's point total in the Leaderboard, which is accessible from the Main Page; in our experiments, users receive 10 points for successful "easy" tasks (bin-picking) and 20 points for successful "difficult" tasks (the remaining tasks). From the Main Page, users can also choose to provide feedback, or press a Request Help button which immediately notifies the study team (e.g., if the user needs assistance or if the setup requires maintenance).

B. Interactive Tutorial

We provide a zoomed-in version of the pages in the Interactive Tutorial in Fig. 13. The aim of the tutorial is to guide the user on how to start and stop interaction episodes as well as how to puppeteer with ALOHA. Specifically, users are first instructed to wait until ALOHA's arms rise to the home position, and then they are given instructions on how to start puppeteering (by squeezing both sets of grippers on the leader arms). After they do so, the tutorial automatically proceeds to the next stage, where users then are told to gently touch the left and right arms to the table; the goal is to help users get calibrated to the robot's range of motion and degrees of freedom, as well as the types of forces they need to apply to move the arms. Finally, users are given instructions on how to stop the interaction episode, by resting the grippers of the leader arms in the grooves of the mechanical stops. When the user does so, the puppet arms are automatically lowered, and the user is presented a brief video on how to navigate the rest of the interface.

C. Implementation Details

The software application is implemented with React (frontend) and Flask (backend), and uses WebSocket connections to communicate between the user client and backend server. We use a SocketIO-ROS bridge to pass messages between the backend server and robot controller. The robot controller operates at 50Hz and is based on [1]. When the robot is being teleoperated, we run a parallel simulation in MuJoCo [52] which is updated at every time step to detect self-collisions.

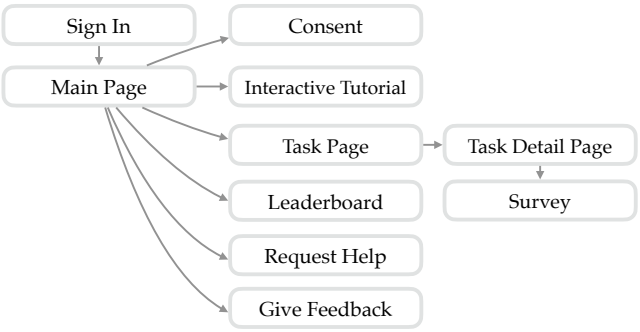


Fig. 12: Flowchart illustration of pages in the user interface.

D. Data Annotation Pipeline

We annotate episodes in our crowdsourced dataset by task and quality. We implement an interface for annotation, which we illustrate in Fig. 15. We annotate episodes by dragging a slider which scrubs through the episode and selecting a task and quality annotation for different segments of the episode. We describe the annotation rules below.

- `play` (Quality 0). All free-play behavior is marked as `play` with quality 0. Play data includes undirected movements and tasks that the user makes up (e.g., trying to unwrap a candy). It also includes extraneous movements before and after the user performs a task.
- `tutorial` (Quality 1–2). Movements associated with the tutorial (e.g., touching the grippers to the table) are marked as Q1 if there is any retrying behavior and Q2 if the motions are smooth.
- `<task>` (Quality 1–3). Task-relevant motions for each of our six tasks are labeled with the task name and a quality from 1 to 3. Q3 is used to describe segments that complete subtasks smoothly with no more than 2 retries. Q2 is used to describe segments that use no more than 4 retries for any one subtask, or that are completed but with slight errors (e.g., grabbing more than 1 candy from a bin). Q1 is used to describe segments that are task-relevant but of poor quality (e.g., more than 4 retries for any one subtask), cause changes to the scene (e.g., dropping a candy on the table), or complete the task in a significantly different manner than the expert demonstrations (e.g., using the opposite arm for any subtask).

Fig. 13: Screenshot of the pages in the interactive tutorial interface.

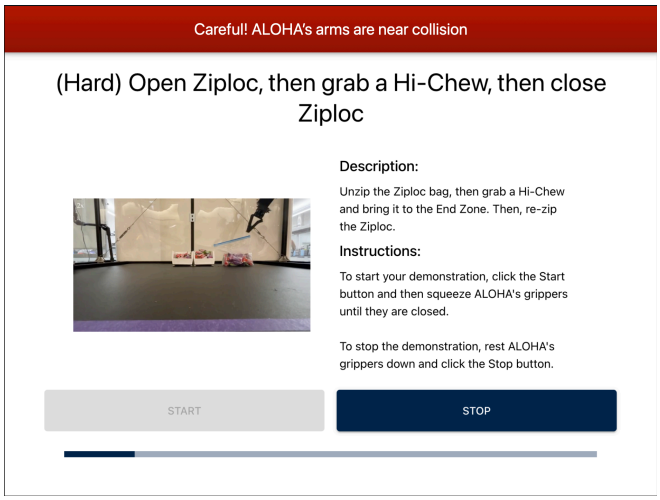


Fig. 14: Screenshot of a visual collision warning on the task page. An audial alarm (beeping sound) is played on the tablet when the visual collision warning appears.

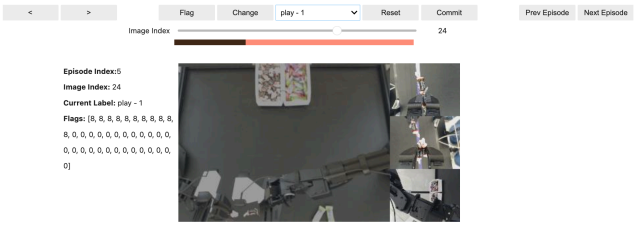


Fig. 15: Screenshot of the data annotation interface. Annotators can scrub through the episode and label segments with task and quality labels, which color codes a bar to visualize the different tasks and qualities in the episode. When the annotator is done labeling an episode, they can "commit" their labels and proceed to the next episode.

APPENDIX VI
ADDITIONAL DETAILS ON PILOT STUDIES AND SYSTEM DEVELOPMENT

Prior to full system deployment, we conducted pilot studies on a smaller population to help us iterate on our system. We obtained the Institutional Review Board's approval before both the pilot studies and the full deployment. We recruited N=10 participants to interact with the system. In order to mimic organic interactions as closely as possible, we did not provide the participants with any verbal instructions, other than to begin interacting with the system as if they happened upon it organically. Our software interface guided the participants through the consent form and tutorial. Here is a sample of feedback provided by participants, coupled with changes we made to the system.

- *Degrees of Freedom*: Users indicated that puppeteering demonstrations was challenging the first time because they needed to "understand the degrees of freedom" of the robot. To address this feedback, we created a tutorial where the user was guided through how to perform primitive movements of the leader arms (e.g., controlling both puppet arms to touch the bottom of the workspace) before they began interacting with the system.
- *Tutorial Format*: In an initial prototype, our tutorial was a

video that a user would watch before using the system. Users provided feedback that they felt "impatient" and would rather "explore what it is like to interface with the robot" rather than "watch a long video." To address this feedback, we made the tutorial efficient and interactive: 4 steps that the user would perform with the robot after watching them on the screen. The interactive tutorial automatically advances after detecting that each step is complete.

- *Start and Stopping Demonstrations*: In an initial prototype, users begin demonstrations by (1) tapping a Start button on an interface and (2) squeezing the grippers of the leader arms closed. To terminate episodes, they would simply need to (1) leave the arms to rest on the robot body and (2) tap a Stop button on the interface. We received feedback that squeezing the gripper to start episodes "made sense" but the "rest position at the end was confusing." To address this feedback, we designed and 3D printed a mechanical stop for users to rest the arms. We automatically terminate episodes when handles of the leader arms make contact with this mechanical stop.
- *Interface*: In an initial prototype, users would access the interface on their own smartphone by scanning a QR code pasted on the platform. A user reported that they would prefer if more of their interaction would happen "in the position that they will be doing the task." We therefore switched to a tablet interface mounted at the base of the platform, which was accessible when the user sat down to begin interacting with the robot. On the interface itself, users reported that it was "easy to understand."
- *Collisions*: We observed that participants did not actively pay much attention to collisions between the robots, as well as the collision of wrist-camera mounts and objects mounted on the table. To address this, we (1) added collision avoidance between the arms and the table, (2) added an audio-visual alarm when arms were near collision, and (3) mounted objects to the table so that they would not move.

APPENDIX VII
OVERVIEW OF ACTION CHUNKING WITH TRANSFORMERS (ACT)

In this section, we provide a more extended background overview of imitation learning (IL) and the Action Chunking with Transformers (ACT) algorithm [1].

Imitation learning (IL) aims to learn a policy π_θ parameterized by θ given access to a dataset $\mathcal{D}$ composed of expert demonstrations. Defined within the framework of a standard partially observable Markov decision process (POMDP), each trajectory $\xi \in \mathcal{D}$ is a sequence of observation-action transitions $\{(o_0, a_0), \ldots, (o_T, a_T)\}$. Most commonly, IL is instantiated as behavior cloning, which trains π_θ to minimize the negative log-likelihood of data, $\mathcal{L}(\theta) = -\mathbb{E}_{(o,a)\sim\mathcal{D}}[\log \pi_\theta(a|o)]$.

In practice, the human-collected demonstrations in $\mathcal{D}$ may be diverse. To effectively learn from such diverse data, we can condition the policy on a latent variable z, which helps to capture the variability in the demonstrations by representing different modes of behavior. Representing this policy as the decoder in a conditional variational autoencoder (cVAE), we in addition learn an encoder q_ϕ from (observation, action) pairs to the latent space: $q_\phi(z \mid a^t, o^t)$. And we condition our policy on the latent variable: $\pi_\theta(\hat{a}^t \mid o_t, z)$. At test time, we sample latent vectors from the standard normal distribution, $z \sim \mathcal{N}(0,1)$. We regularize the outputs of our encoder towards this distribution via a KL-penalty: $D_{KL}(q_\phi(z \mid a^t, o^t) \,\|\, \mathcal{N}(0,1))$. This method is formalized as Action Chunking with Transformers (ACT) [1], an imitation learning algorithm designed to learn from diverse human demonstrations.

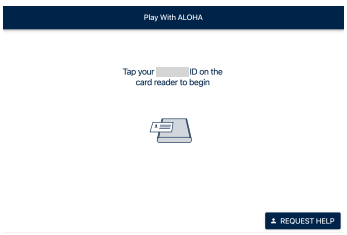
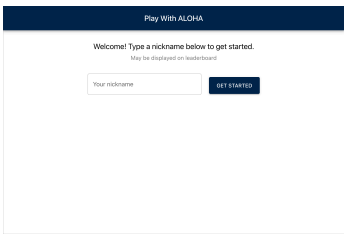
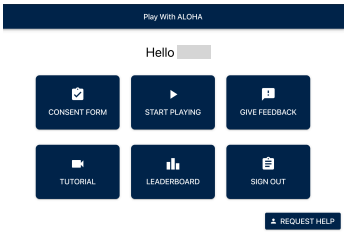
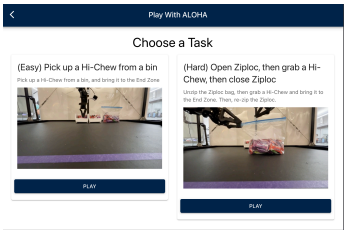
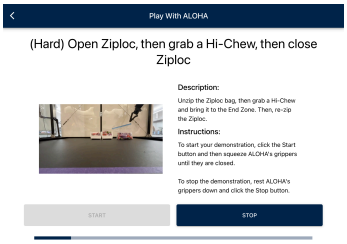
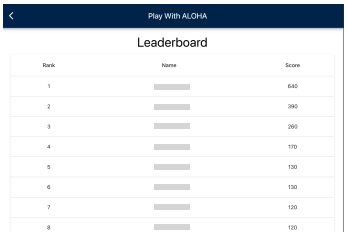
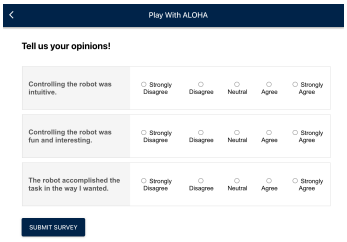
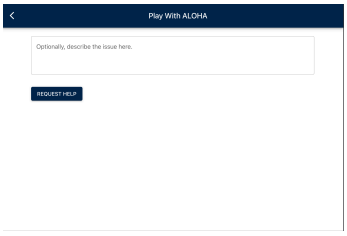
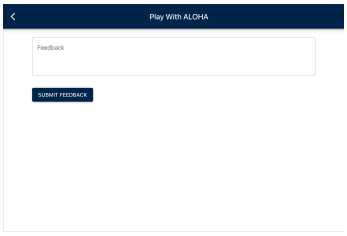
Page Name	Screenshot	Page Name	Screenshot
Sign In (Tap ID Card)		Sign In (Create User Profile)	
Main Page		Interactive Tutorial	
Task Page		Task Detail Page	
Leaderboard		Survey Page	
Request Help		Give Feedback	

TABLE X: Screenshots of pages in the user interface.