Demonstration Sidetracks: Categorizing Systematic Non-Optimality in Human Demonstrations

Shijie Fang[1†], Hang Yu[1†], Qidi Fang[1], Reuben M. Aronson[1], Elaine Schaertl Short[1]

***Abstract*— Prior work in learning from demonstration (LfD) often characterizes the sub-optimalities in human demonstrations as random noise. In this paper, we explored non-optimal behaviors in non-expert demonstrations and showed that these behaviors are not random and follow systematic patterns: they form systematic demonstration sidetracks. We used a public space study dataset from our previous work with 40 participants and a long-horizon robot task. We recreated the experimental setup in a simulation and annotated all the demonstrations. We identified five types of demonstration sidetracks, *Exploration*, *Mistake*, *Alignment*, *Pause*, and *One-dimension control*. We found that instead of being random and rare, demonstration sidetracks frequently appear in non-expert demonstrations across all participants, and the distribution of demonstration sidetracks correlates with the robot task temporarily and spatially. Moreover, we found that users' control patterns are associated with the control interface. Our findings highlight the need for better models of non-expert behavior, offering insights to improve LfD algorithms and reduce gaps between lab-based training and real-world applications. All demonstrations we used, infrastructures for annotation, and annotation results are available at https://github.com/AABL-Lab/Human-Demonstration-Sidetracks**

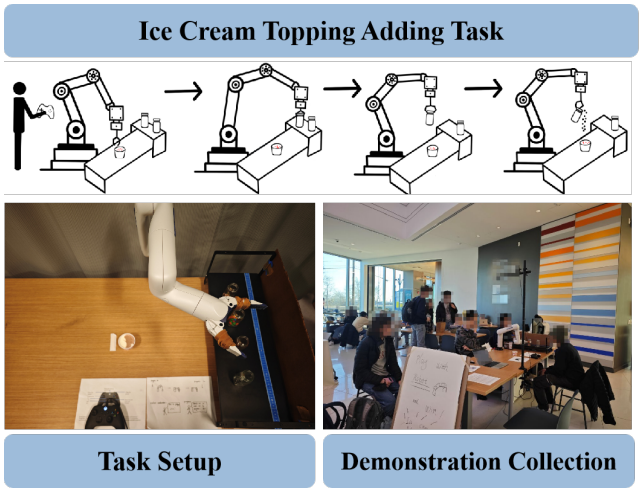

Fig. 1. Public space study and ice cream topping adding task. Participants provided demonstrations in a non-lab setup with a long-horizon task. Participants controlled the robot arm to first pick up a top jar, and then pour the toppings onto an ice cream.

I. Introduction

Learning from human demonstration (LfD) has emerged as a crucial technique for robot learning [1], and has been successful in many contexts [2] [3] [4] [5], ranging from single-step to long-horizon tasks. More recent work has demonstrated that using non-expert demonstrations could reduce the data needs for expert demonstrations if the noise in non-expert demonstrations can be mitigated. Much prior work assumes that non-expert demonstrations are low-quality and noisy [6]. However, there is little research on understanding and modeling non-expert demonstrations, which are important for developing and testing LfD algorithms. Our key insights are that: *not all non-optimal behaviors from non-experts are random: instead they appear in structured ways.* Thus, in this work, we focus on investigating patterns and meanings of the noise in non-expert demonstrations.

Modeling non-expert behaviors is important for learning from non-expert demonstrations, and can be beneficial for both learning and validation. For instance, one important technique to enable learning from non-expert demonstrations is modeling non-expert demonstrations explicitly and designing learning algorithms to be able to adapt to these noisy or sub-optimal demonstrations. Oracles are used in prior work to provide non-expert demonstrations to validate algorithm performance, and non-expert behaviors are modeled by injecting noise into expert trajectories generated by machine-learning methods. [7]

Inaccurate modeling of non-expert behaviors can result in a performance gap between experiments and real-world scenarios. Methods that work with oracle demonstrations might not perform well with real human demonstrations. However, few works have studied patterns in non-experts' behavior when giving robot demonstrations. Prior work often describes non-expert demonstrations using relatively simple patterns, such as expert demonstrations with a certain amount of noisy actions [7], or expert policy mixing with random policies [8], which failed to capture the diverse strategies or accommodation techniques non-experts use to succeed.

Our goal is to investigate non-expert behaviors, and to explore the patterns of noise in non-expert demonstrations. In this work, we analyzed data from a public space study with 40 non-expert demonstrators with a long-horizon task, from our previous work initially introduced to test a novel feedback mechanism [9]. We closely defined the frequent and patterned non-optimal behaviors that non-experts performed during demonstrating as *demonstration sidetracks*. To identify demonstration sidetracks in the collected demonstrations, we designed a graphical interface and an open-coded book to annotate the demonstrations, and cross-validated the labels with a third expert. All codes and data we annotated are

†These authors contributed equally to this work.
[1]Tufts University School of Engineering, Computer Science. Medford, Massachusetts, United States of America {shijie.fang, hang.yu625917, qidi.fang, reuben.aronson, elaine.short}@tufts.edu

published online. We found that demonstration sidetracks frequently exists in non-expert demonstrations across all demonstrators. We also show that demonstration sidetracks are not randomly distributed in demonstrations. When and where demonstration sidetracks appeared have temporal and spatial relationships with the task. The demonstration sidetracks are more frequently identified when the sub-goal has changed, or accurate manipulation or perceptions are required. Moreover, user behavior patterns are also associated with the control interface.

The main contributions of this work are characterizing and categorizing non-optimal behaviors in human demonstrations. We found non-optimal behaviors are beyond random errors, and possess task-related structures. To the best of our knowledge, this is the first work that systematically investigates and characterizes non-optimal manipulation patterns from non-experts when providing demonstrations for robot tasks. We believe that our work will be beneficial for future LfD algorithm development to better learn from noisy demonstrations and oracle design to reduce the gap between in-lab datasets and real-world demonstrations.

II. Background

A wide variety of learning from demonstration methods has emerged in the past few decades. Overcoming and understanding imperfections in human demonstration is crucial for the robustness of LfD methods, especially in real-world scenarios [10].

A. Learning from demonstrations

Learning from demonstrations allows robots to learn a task policy from demonstrations provided by human teachers [11]. Behavioral Cloning (BC) is a straightforward way of learning from demonstration [12]. BC is capable of learning complex robot behaviors and performant general robot policies with sufficient and numerous demonstrations [13] [14] [2]. Inverse reinforcement learning (IRL) views the LfD process as recovering reward functions from human demonstrations and learning a robot policy via reinforcement learning from inferred reward functions [15] [16]. Other methods drawing intuition from IRL inherit its basic idea but vary in terms of network structures. [17] [18] Most regular LfD methods require high-quality demonstrations to work [10]. Difficulty in obtaining sufficient expert demonstrations forms a major challenge for LfD methods. As a result, efficiently learning from non-expert demonstrations is crucial to the wide deployment of LfD-based algorithms. This addresses the need to look into actual patterns in non-expert demonstrations.

B. Learning from non-expert demonstrations

Enabling LfD to learn from sub-optimal or non-expert demonstrations is a promising approach to lifting data availability and robustness for robot learning [19]. Learning methods need to infer the distribution of the optimal policy from noisy or sub-optimal distributions. One applicable method is by introducing extra human knowledge, such as ranking [20], preference [21], feedback [22], and confidence [23], [24]. Other methods introduce prior knowledge about the possible distribution of optimal policies. This prior knowledge can come from a small set of experts or optimal demonstrations [10] [25], or pre-designed metrics about the task [26]. Prior knowledge allows LfD algorithms to implicitly rank or weight sub-optimal demonstrations based on their distance to given optimal demonstrations or desired metric values, resulting in a better inference about the true optimal policy. However, including extra human knowledge can be laborious and time-consuming. Also, this extra information is typically task-specific so it can not be used "off-the-shelf" to enable learning from non-expert demonstrations in new scenarios.

C. Modeling Noisy Demonstrations

In order to learn from non-expert demonstrations without introducing extra human feedback or prior knowledge, modeling sub-optimality becomes necessary for learning from sub-optimal demonstrations. The modeling of sub-optimality includes the modeling of the type of distribution [27], scale [7], and their relationships to other variables [6]. With the modeling of sub-optimal demonstrations, optimal policies can be recovered through selective matching [7], denoising [28], automatic ranking and weighting [6]. Despite increasing focus on learning from imperfect human demonstrations, work that uses real-world imperfect human demonstrations is still relatively rare. Prior work generally has relatively simple assumptions on sub-optimality modeling: like being a mixed Gaussian distribution of optimal and noise [8] [27] [6] or having optimal the part being dominant while sub-optimality only takes up a small part [7].

Our work differs from prior work by categorizing and identifying patterns in non-optimal robot demonstrations, and acknowledging some non-optimal behaviors are natural for humans. The demonstrations we used in this work were collected outside the lab with random people. Our data and results can be expected to better reflect robot-in-the-wild deployments than in-lab setups.

III. Methodology

Giving demonstrations to a robot is a complex and high-demanding task. Unlike well-trained agents, humans exhibit sub-optimal behaviors while manipulating robots. However, not all imperfect behaviors are meaningless or errors. We believe that some non-optimal behaviors are beyond random noise or mistakes, they are actually performed by the demonstrators intentionally or subconsciously to provide successful demonstrations. In this work, we conducted an exploratory analysis of a demonstration dataset to investigate noisy behaviors in non-expert human demonstrations. **Our key insights** are that:

> *Non-optimal behaviors in human demonstrations are structured with patterns.*

The demonstrations came from 40 non-expert demonstrators who were asked to provide demonstrations over a long horizon task. In this work, we annotated all the demonstrations using an open-coded book with a graphical interface.

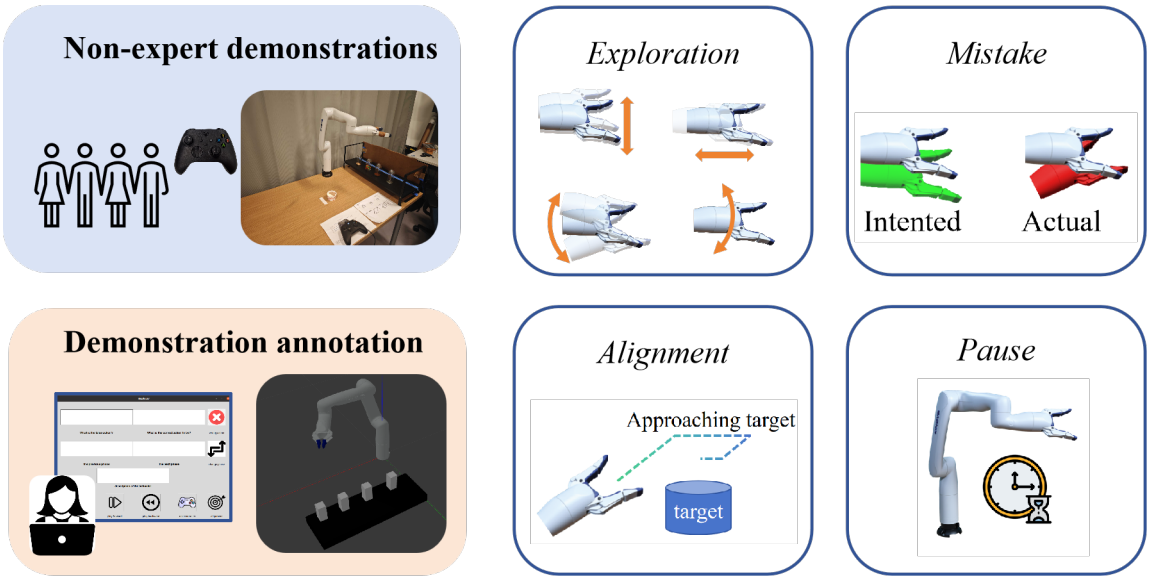

Fig. 2. Experiment pipeline and demonstration sidetracks. Non-expert demonstrations were collected by having participants control a real robot. We then replayed demonstrations in simulation to annotate the demonstration sidetracks. We identified five types of demonstration sidetracks and illustrated four of them – *Exploration*, *Mistake*, *Alignment*, and *Pause* – on the right side of our figure.

A. Demonstration Sidetracks

Prior work interprets non-optimal behaviors in human demonstrations as random manipulation noise or incorrect behaviors with underlying optimal control policies [27] [6] [8]. However, unlike trained agents or oracles, which can control all action dimensions instantly and are either provided with accurate observation or trained to reduce perception error, humans are limited by the control interface, reaction time, and inaccurate perceptions.

Intuitively, in order to complete the task and overcome these limitations while maintaining a reasonable mental workload, humans can accommodate these limitations by controlling a limited degree of freedom at a time, performing redundant but harmless behaviors, and spending more time while giving demonstrations. Unlike random noise, these behaviors would appear consistently within both the overall policy and concrete control behaviors, and would have special task-related temporal and spatial characteristics. As a result, human demonstrations tend to be different from oracle demonstrations by containing not only noisy actions but also accommodation behaviors. In this work, we define these behaviors in human demonstrations as *Demonstration Sidetracks*.

B. Experiment Setup

The demonstration data analyzed in this work was collected as part of a validation study for a novel teaching signal [9]. We review the study design here, which is also described in [9]. The robot used for the data collection was a Kinova Gen3 Lite robot with six degrees of freedom. Participants used an Xbox controller to control the robot arm. The task used for collecting human demonstrations is an ice cream topping-adding task.

The setup and the control interface are shown in Figure 2. The task consisted of six sub-tasks: move down to jar, approach jar, grasp jar, lift jar, move to ice cream cup, approach ice cream cup, and pour. Participants first control the robot to reach a jar they selected on a shelf, pick up the jar, navigate to the ice cream location, and then pour the toppings into the ice cream cup. This task is suitable for our study because it features both simplicity and diversity of primal sub-tasks. The overall task is relatively simple and has an appropriate length for manual annotation. It includes diverse sub-tasks and primal actions that cover some of the most common robot skills, including free-space reaching, approaching small objects, grasping, aligning the robot to the target position, and object pose manipulation [5].

All control actions were in end-effector space, and actual joint motions were calculated in real-time using forward kinematics. The end-effector movements were decomposed into several primitive movements, including translations along the X, Y, and Z axes, rotations in roll, pitch, and yaw, as well as opening and closing of the gripper.

C. Experiment procedure

The public space study was held in the atrium of a university building to collect non-expert demonstrations in a non-lab setup, as shown in Figure 1. Participants were recruited from random people walking past the experiment setup. 40 participants were recruited in total. 22 were male, 14 were female, and 4 preferred not to say. Participants were asked to provide one demonstration and had up to three minutes to practice the task. Participants only had one shot to demonstrate the task, and could not retry if they failed. The demonstration data were recorded at a frequency of 5Hz in a form of joint positions. More details of the experimental procedures can be found in [9].

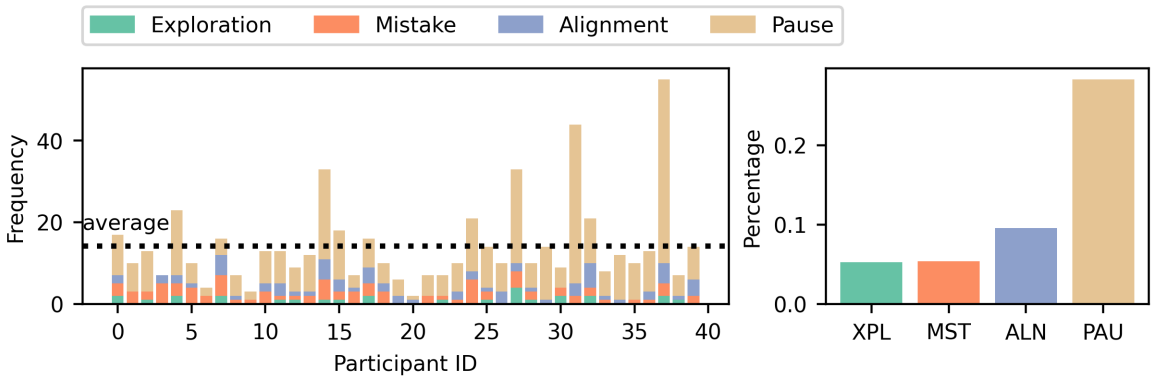

Fig. 3. Demonstration sidetracks frequency and ratio. The left figure shows the number of demonstration sidetracks observed across all demonstrators. The X-axis is participant ID, and the Y-axis is the number of each type of demonstration sidetrack. The figure on the right side shows the percentage of time spent on each type of demonstration sidetracks relative to the total time. The results indicate that demonstration sidetracks widely and frequently exist in non-expert demonstrations.

D. Open-Coded Book

Based on our observation during the study and our intuitions on possible imperfections, we propose and study five types of demonstration sidetracks in this work:

1) Exploration: A sequence of primal movements that include multiple dimensions of freedom and are often irrelative to completing the task.

2) Mistake: A sequence of homogenous movements, meaning only one type of primal movement is included, that has no contributions to, or harms the completion of the task.

3) Pause: Significant pause during the execution of the task. In our case, one second with no control inputs is considered a pause.

4) Alignment: Back-and-forth movements that align the robot with the target object. We used two additional features to help to identify the intention of alignment. The first feature is actual precision increase, meaning the robot is closer after this behavior. The second feature is a switch of sub-tasks, meaning the sub-task will likely change to the next sub-task after alignment.

5) One-dimension control: Tendency of control using a single primal movement at a time.

Here, we define a primal movement as a movement that is mapped to a single input source on the controller interface, i.e., one button or joystick. In the case of our study, the seven primal movements are moving along the X, Y, and Z axes, roll, pitch, yaw, and operating grippers.

E. Replay and annotation

We recreated our experiment setup in a simulated robot environment, and annotated the demonstrations in the simulation. Simulation allows us to control the speed of replaying and pause or play backward at any step, which helps us to catch more subtle non-expert behavior and annotate their precise starting and ending times. All *pauses* were automatically annotated by a script. In addition to demonstration sidetracks, we also annotated task phases to study the temporal relationships between non-expert behaviors and task phases.

We designed a graphical user interface to help with data annotation. The interface includes three buttons for controlling the replay, three buttons for labeling demonstration sidetracks, one button for marking phase change, and five input boxes for additional comments. This interface allows annotators to pause the replay, play forward, play backward, label behaviors and task phases, and add additional descriptions such as start and end steps for the labels.

Two robot experts conducted the overall annotation, which was crosschecked by the third robot expert. Behaviors that appear both in the original annotation and crosscheck annotation are adopted for our data analysis.

IV. Results

We manually annotated exploratory actions by replaying 40 real-world robot demonstrations from non-expert demonstrators in simulation according to our proposed open-coded book. Our main goal for conducting data analysis is to identify demonstration sidetracks and unveil the underlying patterns of these behaviors. We found that demonstration sidetracks widely existed in the demonstrations, demonstration sidetracks more frequently appeared around the target objects or when the sub-task has changed, and demonstrators tended to control one dimension at a time.

A. Commonality in demonstrations

In this subsection, we address the importance of identifying and characterizing demonstration sidetracks by showing that these demonstration sidetracks widely existed in human demonstrations, even in demonstrations that successfully complete the task.

We show the amount of demonstration sidetracks for each participant in Figure 3. We found that demonstration sidetracks widely existed across all demonstrations: while 33 out of 40 non-expert participants successfully demonstrated the task, the average number of demonstration sidetracks is 14.48 ± 10.54 per participant. We calculated the percentage of timesteps from demonstration sidetracks compared to total timesteps recorded. We show timesteps from demonstration

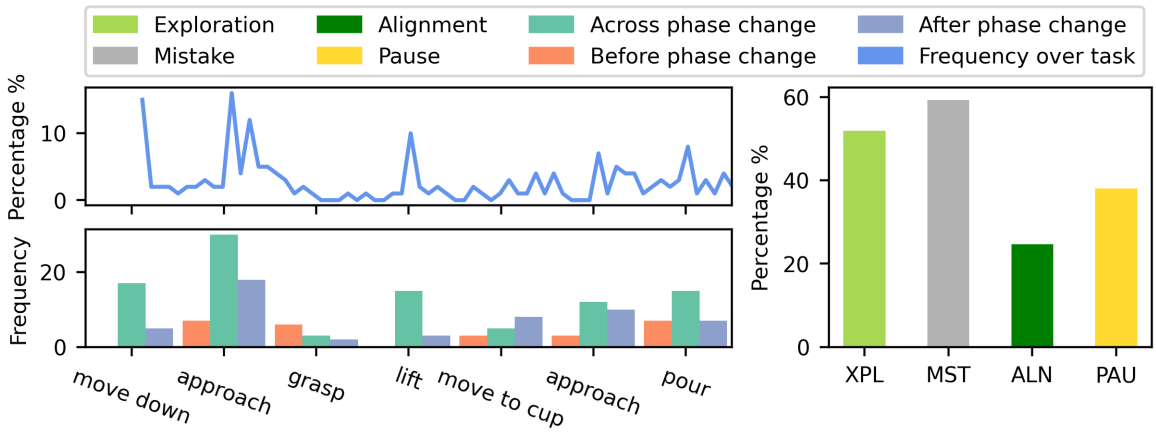

Fig. 4. Temporal relationships between task phase changes and demonstration sidetracks. The upper-left part shows the percentage of demonstration sidetracks in different sub-tasks. The bottom-left part shows the number of demonstration sidetracks happened within 40% percent timesteps window around the phase change. The X-axis displays different sub-tasks in order, and the Y-axises are percentages and frequency separately. The right part shows the percentage of each type of demonstration sidetracks happening within a 4-second window around the task phase change. We found that the occurrence of demonstration sidetracks is associated with the change of the sub-tasks.

sidetracks take up 48% of all timesteps recorded in the overall task. This suggests that, contrary to the common assumption that imperfections only take up a small part in non-expert demonstration, demonstration sidetracks are frequent and take up a large portion of the overall task demonstration. Additionally, we observed a high diversity in demonstration sidetracks types in a single demonstration, with 62.5% demonstrations possessing three or more types of demonstration sidetracks.

B. Demonstration Sidetracks are more likely to happen when sub-task changes

We show the number of demonstration sidetracks in this subsection, we look into the temporal distribution of demonstration sidetracks.

One core assumption for drawing noises from simple distribution is that these noises are not temporally correlated with the task itself. However, we show that this is not the case for actual human demonstrations by analyzing the relevance of demonstration sidetracks with the task's progress.

We analyzed the relative time when demonstration sidetracks occurred during different task phases, shown in Figure 4. We found that demonstration sidetracks were more likely to occur near the moment when the task phase changed. 54.7% of overall demonstration sidetracks behaviors occurred in the 40% range across the instance of the task change. This is likely due to the different sub-goals of different task phases, along with a change in the control pattern for participants, resulting in increasing demonstration sidetracks.

Additionally, we analyzed the absolute time when demonstration sidetracks happened. We analyzed the percentage of different demonstration sidetracks happening within four seconds from the instance of the task phase change. Results show that 52% of *Exploration*, 59% of *Mistake*, and 38% of *Pause* happened within four seconds from task phase change instance, further revealing that humans tend to have demonstration sidetracks near task phase change temporally.

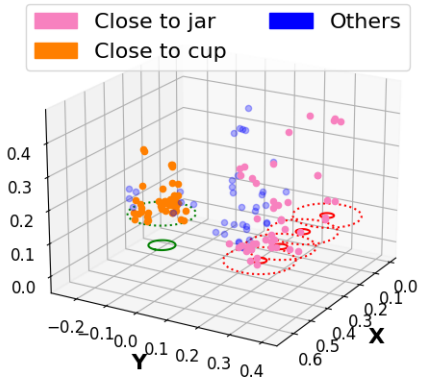

Fig. 5. The spatial distribution of *Alignment* behaviors. The points represent the position of the robot end-effector position at the start and end of each *Alignment* behavior. The red circle represents the positions of four jars, while the green circle represents the ice cream cup. Points within 0.1 meters from the jars and the cup are highlighted in orange and pink. Other points outside this range are illustrated in blue. We found that *Alignment* behaviors more frequently occurred around target objects.

C. Alignment happens more when accuracy is required for controlling robot

Here, we examine the spatial distribution of demonstration sidetracks. Simple random noise is usually sampled step-by-step, but the actual robot location is rarely considered. We show that the robot's location is an important task-related factor affecting human behavior. We reveal that precision requirements affected the spatial distribution of demonstration sidetracks.

We analyzed the relationship between the number of demonstration sidetracks and task phases. We show that *Alignment* occurred more often when the task was to ap-

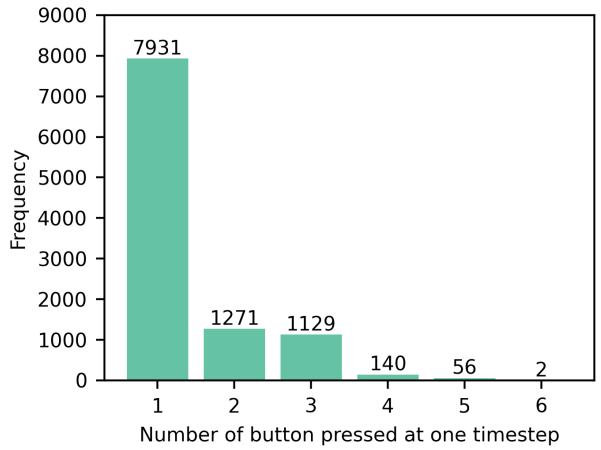

Fig. 6. The number of dimensions controlled at a single timestep. The X-axis shows the number of dimensions controlled at a single recorded timestep, while the Y-axis shows the frequency. Results show that participants only controlled one dimension, i.e. one button or one joy stick was pushed, for 75% of the time.

proach a certain target position, as shown in Figure 5. 73% of all *Alignment* behaviors started and ended within 0.1-meter range from the target. Among these, 40% *Alignment* behaviors started and ended near the jars, and 33% of all alignment behaviors happened near the ice cream cup, which was the target for the pouring task.

This suggests that demonstration sidetracks, especially *Alignment*, are more likely to happen when approaching a small target object. This is possibly due to the requirement for control accuracy, which resulted in more alignment to get to the desired positions and more pauses for reaction.

D. Effect of Control Interface

In addition to the imperfect manipulation that leads to a change in the local control pattern, we also found that the control interface affected participants' overall manipulation strategies. We found that there was a common tendency to control only one dimension at a time in their manipulation patterns. We analyzed the number of buttons pushed during each record timestep by measuring the position change of the robot end effector before and after each recording timestep and extracted the number of action taken. Results show that 75% of all timesteps when the robot was moving only recorded a single action, whereas 12% recorded two actions and 10% recorded three actions, as shown in Figure 6. This tendency of controlling fewer dimensions simultaneously can cause human policies to deviate significantly from the optimal policy.

V. Discussion

In this work, we found that human imperfections when giving demonstrations to robots are not just simple noise or randomly distributed. Rather, these imperfections usually have distinct patterns, are often intentional, and are affected by the control interface, which causes real-world human demonstrations to differ from oracle demonstrations. We defined these meaningful or patterned imperfections as demonstration sidetracks.

We believe the differences between oracle demonstrations and human demonstrations lie in the following factors. **Human control of robots is limited by the input interface.** For instance, in our experiment, participants controlled the robot using an Xbox controller, and each button or joystick controlled one dimension of freedom. Controlling movements in multiple dimensions requires a higher mental load, and thus was less preferred even if it was more optimal. **Humans need reaction time when controlling robots**, which could result in frequent pauses. Lastly, unlike oracles, which are provided with unambiguous observations and trained to have consistent reactions, **human perceptions and controls are less precise**. To succeed in the task, demonstrators may perform a sequence of sub-optimal actions instead of one optimal action to mitigate errors.

One limitation of this work is that we only used demonstrations from one task. Even though the task has covered a good range of sub-tasks (e.g. picking and pouring), the work will benefit from non-manipulation tasks. For future works, we will develop quantitative metrics for demonstration sidetracks and develop methods that could identify demonstration sidetracks automatically. Another future work is to design a method that injects demonstration sidetracks into synthetic oracle demonstrations to generate more realistic demonstrations for fast algorithm development.

VI. Conclusion

In this work, we explored systematic human non-optimal behaviors, defined as *Demonstration Sidetracks*, in a real-world robot demonstration task. We identified five categories of demonstration sidetracks from the demonstrations. We show that demonstration sidetracks are frequent, correlated with the task both temporally and spatially, and have effects on human control policies. Our findings highlight the need for more realistic modeling of sub-optimality in human demonstrations. We believe this work lays a foundation for bridging the gap between laboratory-based setups and real-world scenarios.

References

[1] H. Ravichandar, A. S. Polydoros, S. Chernova, and A. Billard, "Recent advances in robot learning from demonstration," *Annual review of control, robotics, and autonomous systems*, vol. 3, no. 1, pp. 297–330, 2020.

[2] O. M. Team, D. Ghosh, H. Walke, K. Pertsch, K. Black, O. Mees, S. Dasari, J. Hejna, T. Kreiman, C. Xu *et al.*, "Octo: An open-source generalist robot policy," *arXiv preprint arXiv:2405.12213*, 2024.

[3] E. Jang, A. Irpan, M. Khansari, D. Kappler, F. Ebert, C. Lynch, S. Levine, and C. Finn, "Bc-z: Zero-shot task generalization with robotic imitation learning," in *Conference on Robot Learning*. PMLR, 2022, pp. 991–1002.

[4] E. Johns, "Coarse-to-fine imitation learning: Robot manipulation from a single demonstration," in *2021 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2021, pp. 4613–4619.

[5] A. Padalkar, A. Pooley, A. Jain, A. Bewley, A. Herzog, A. Irpan, A. Khazatsky, A. Rai, A. Singh, A. Brohan *et al.*, "Open x-embodiment: Robotic learning datasets and rt-x models," *arXiv preprint arXiv:2310.08864*, 2023.

[6] M. Beliaev, A. Shih, S. Ermon, D. Sadigh, and R. Pedarsani, "Imitation learning by estimating expertise of demonstrators," in *International Conference on Machine Learning*. PMLR, 2022, pp. 1732–1748.

[7] F. Sasaki and R. Yamashina, "Behavioral cloning from noisy demonstrations," in *International Conference on Learning Representations*, 2020.

[8] L. Liu, Z. Tang, L. Li, and D. Luo, "Robust imitation learning from corrupted demonstrations," *arXiv preprint arXiv:2201.12594*, 2022.

[9] H. Yu, Q. Fang, S. Fang, R. M. Aronson, and E. S. Short, "How much progress did i make? an unexplored human feedback signal for teaching robots," *arXiv preprint arXiv:2407.06459*, 2024.

[10] H. Xu, X. Zhan, H. Yin, and H. Qin, "Discriminator-weighted offline imitation learning from suboptimal demonstrations," in *International Conference on Machine Learning*. PMLR, 2022, pp. 24 725–24 742.

[11] S. Chernova and A. L. Thomaz, *Robot learning from human teachers*. Morgan & Claypool Publishers, 2014.

[12] D. A. Pomerleau, "Alvinn: An autonomous land vehicle in a neural network," *Advances in neural information processing systems*, vol. 1, 1988.

[13] P. Florence, C. Lynch, A. Zeng, O. A. Ramirez, A. Wahid, L. Downs, A. Wong, J. Lee, I. Mordatch, and J. Tompson, "Implicit behavioral cloning," in *Conference on Robot Learning*. PMLR, 2022, pp. 158–168.

[14] C. Chi, S. Feng, Y. Du, Z. Xu, E. Cousineau, B. Burchfiel, and S. Song, "Diffusion policy: Visuomotor policy learning via action diffusion," *arXiv preprint arXiv:2303.04137*, 2023.

[15] P. Abbeel and A. Y. Ng, "Apprenticeship learning via inverse reinforcement learning," in *Proceedings of the twenty-first international conference on Machine learning*, 2004, p. 1.

[16] B. D. Ziebart, A. L. Maas, J. A. Bagnell, A. K. Dey *et al.*, "Maximum entropy inverse reinforcement learning." in *Aaai*, vol. 8. Chicago, IL, USA, 2008, pp. 1433–1438.

[17] J. Ho and S. Ermon, "Generative adversarial imitation learning," *Advances in neural information processing systems*, vol. 29, 2016.

[18] J. Fu, K. Luo, and S. Levine, "Learning robust rewards with adversarial inverse reinforcement learning," *arXiv preprint arXiv:1710.11248*, 2017.

[19] Q. Wang, R. McCarthy, D. C. Bulens, F. R. Sanchez, K. McGuinness, N. E. O'Connor, and S. J. Redmond, "Identifying expert behavior in offline training datasets improves behavioral cloning of robotic manipulation policies," *IEEE Robotics and Automation Letters*, 2023.

[20] D. Brown, W. Goo, P. Nagarajan, and S. Niekum, "Extrapolating beyond suboptimal demonstrations via inverse reinforcement learning from observations," in *International conference on machine learning*. PMLR, 2019, pp. 783–792.

[21] S. Kuhar, S. Cheng, S. Chopra, M. Bronars, and D. Xu, "Learning to discern: Imitating heterogeneous human demonstrations with preference and representation learning," in *Conference on Robot Learning*. PMLR, 2023, pp. 1437–1449.

[22] J. Huang, R. M. Aronson, and E. S. Short, "Modeling variation in human feedback with user inputs: An exploratory methodology," in *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '24. New York, NY, USA: Association for Computing Machinery, 2024, p. 303–312. [Online]. Available: https://doi.org/10.1145/3610977.3634925

[23] S. Zhang, Z. Cao, D. Sadigh, and Y. Sui, "Confidence-aware imitation learning from demonstrations with varying optimality," *Advances in Neural Information Processing Systems*, vol. 34, pp. 12 340–12 350, 2021.

[24] Y.-H. Wu, N. Charoenphakdee, H. Bao, V. Tangkaratt, and M. Sugiyama, "Imitation learning from imperfect demonstration," in *International Conference on Machine Learning*. PMLR, 2019, pp. 6818–6827.

[25] L. Yu, T. Yu, J. Song, W. Neiswanger, and S. Ermon, "Offline imitation learning with suboptimal demonstrations via relaxed distribution matching," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, no. 9, 2023, pp. 11 016–11 024.

[26] X. Bu, W. Li, Z. Liu, Z. Ma, and P. Huang, "Aligning human intent from imperfect demonstrations with confidence-based inverse soft-q learning," *IEEE Robotics and Automation Letters*, 2024.

[27] M. Beliaev and R. Pedarsani, "Inverse reinforcement learning by estimating expertise of demonstrators," *arXiv preprint arXiv:2402.01886*, 2024.

[28] Y. Yuan, X. Li, Y. Heng, L. Zhang, and M. Wang, "Good better best: Self-motivated imitation learning for noisy demonstrations," *arXiv preprint arXiv:2310.15815*, 2023.