Effects of Integrated Intent Recognition and Communication on Human-Robot Collaboration

Mai Lee Chang[1], Reymundo A. Gutierrez[2], Priyanka Khante[1], Elaine Schaertl Short[1], Andrea Lockerd Thomaz[1]

Abstract— **Human-robot interaction research to date has investigated intent recognition and communication separately. In this paper, we explore the effects of integrating both the robot's ability to generate intentional motion and predict the human's motion in a collaborative physical task. We implemented an intent recognition system to recognize the human partner's hand motion intent and a motion planner system to enable the robot to communicate its intent by using legible and predictable motion. We tested this bi-directional intent system in a 2-way within-subjects user study. Results suggest that an integrated intent recognition and communication system may facilitate more collaborative behavior among team members.**

I. INTRODUCTION

Successful social robot teammates deployed for the long term will need the capability to reason about human intentions as well as communicate their own intentions. Understanding intent includes recognizing the current activity, inferring the task goal, and predicting future actions. Humans easily infer and communicate intent and this capability is particularly critical during collaborative activities. For instance, playing team sports, assembling furniture together, and preparing a meal together all require seamless coordination among the collaborators in which good predictions of others' future actions as well as displaying transparent intentional behavior are both important. In the assembling furniture scenario, for example, as soon as one person starts handing over a tool, the other person knows to reach for it. In a cooking scenario, both collaborators may need to add ingredients at the same time. The dynamics of the interaction changes rapidly between intent recognition and communication. Thus part of being a good collaborator is anticipating other's needs and responding appropriately as well as communicating its own needs. In our work, we show that a robot that is able to both recognize and communicate intent results in better overall team performance.

Prior work in human-robot interaction (HRI) research to date has only investigated intent recognition and communication separately. In this work, we take a holistic approach by exploring the effects of both intent recognition and communication in a human-robot collaborative cup pouring task. We evaluate our work in a scenario where a robot and human collaborate need to pick up two different cups and pour them into the same target container, and where the human gets first choice of cup and the robot chooses the target. In order for the robot to select a different cup from the human, it uses intent recognition to infer which cup the human intends to pick up. The robot then uses its arm motion trajectory to communicate which container it intends to empty the cup into, so that the human can infer the robot's intention and pour into the same container. We implemented an integrated system that consist of an intent recognition system to enable the robot to recognize the human's hand motion intent and a motion planner system to enable the robot to communicate its intent by displaying legible and predictable motion.

We conduct a 2-way within-subjects user study to evaluate the integrated system in four conditions (Figure 1). We show that when the robot recognizes and communicates intent it results in more collaborative behavior in the team.

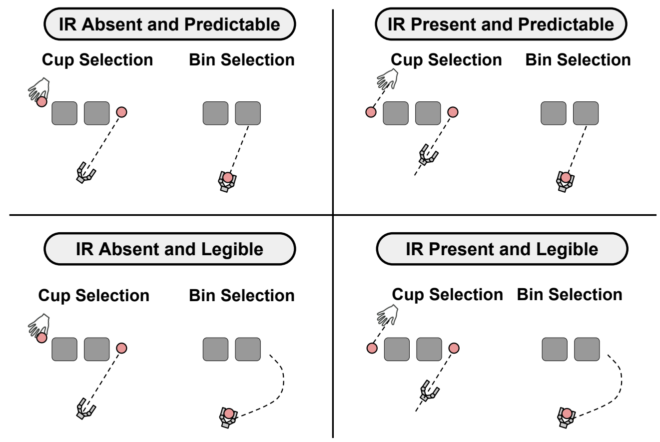


Fig. 1. The task utilized in this study consisted of collaboratively emptying cups into bins. There are four conditions. In the conditions with IR Absent, intent is detected after the cup is grasped whereas with IR Present, intent is detected prior to grasping. With Predictable Motion, bin selection inference is delayed until the gripper is close to the bin whereas with Legible Motion, the inference occurred earlier.

II. RELATED WORK

In the HRI domain, prior work covered two categories: intent recognition and intent communication. For intent recognition, Hoare and Parker [1] used Conditional Random

This material is based upon work supported by the Office of Naval Research award numbers N000141612835, N000141612785 and National Science Foundation award numbers 1564080, 1724157.

[1]Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78705, USA mlchang@utexas.edu, priyanka.khante@utexas.edu, elaine.short@utexas.edu, athomaz@ece.utexas.edu

[2]Department of Computer Science, University of Texas at Austin, Austin, TX 78705, USA ragtz@cs.utexas.edu

Fields to classify the human's intended goal in a box pushing task. Another method used object affordances to anticipate the human's next activity in order to enable the robot to plan ahead for a reactive response [2]. Mainprice and Berenson [3] developed a framework where the motion planner takes into account an estimation of the workspace the human will occupy and showed that this leads to safer and more efficient team performance. Other research also explored how to enable robots to anticipate collaborative actions in the presence of uncertain sensing and task ambiguity. One approach consisted of using probabilistic graphical model of the structured tasks to allow the robot to appropriately time its actions [4]. Another approach utilized anticipatory knowledge of the human motions and subsequent actions to predict the human's reaching motion goal in real-time [5]. Gaze patterns are used to predict the human's intent to achieve efficient human-robot collaboration [6].

Besides understanding intent, robots must be able to also communicate their own intentions. Prior HRI work is inspired by animation techniques and focused on designing human-like robot behavior so that they are intuitive to understand [7]. For instance, different levels of a robot's exaggerated motion is perceived differently by the human collaborator and also affected the retention of the interaction details [8]. Gielniak and Thomaz showed that the spatiotemporal correspondence of actuators can be used to generate motions that better convey intent [9], [10]. Besides manipulation motions, navigation motions including free-flyer robots followed a similar approach to communicate path intentionality [11].

Dragan et al. introduced an approach that takes into account an observer into motion planning [12]. This resulted in motions that are predictable (i.e., motion that matches what the collaborator would expect to see given a goal) and legible (i.e., motion that expresses the robot's intent and allows the collaborator to quickly and confidently identify the goal). They showed that legible motion resulted in more fluid collaborations as in comparison to predictable motion [13]. The intent communication aspect of our work is based on these results. Our work took a holistic approach and investigated intention as a bi-directional interaction. We implemented an integrated intent recognition and communication system and investigated its effects on the human-robot team performance.

III. INTENT RECOGNITION

In this work, we infer human intent by recognizing whether the person is reaching to the left/right of the workspace. There are two intent recognition systems compared in this study: object intent recognition and hand motion intent recognition. We hypothesize that the hand motion intent system would enable the robot to predict the human's intent faster as in comparison to the object intent recognition. Since our task is a collaborative cup pouring task where the human gets to select a cup first, the hand motion intent system would detect which cup the human selects before the human grabs the cup. On the other hand, the object intent

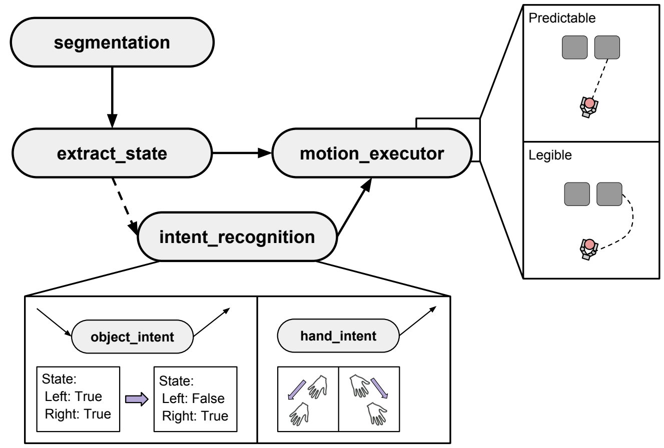


Fig. 2. The integrated intent recognition and generation system consists of four modules. The intent recognition and motion execution modules are set according to the experimental conditions.

recognition system would predict the selection of the cup later, i.e., after the human has already grabbed the cup. These are the two intent recognition systems (absent vs. present) that we will later use in our user study.

At a high level, the object intent recognition system detects intent through objects being moved in the scene. In order to accomplish this, we use planar segmentation to identify the cups on the table and extract the presence/absence state of each cup. At each time step, the object intent recognition module looks for changes between the current and previous state. If a change is detected, this module outputs the side that has an object removed as the human's intent. This provides a baseline intent recognition system, as the robot only recognizes intent after an object has been moved.

The hand motion intent recognition system detects human intent by tracking the motion of a blue-colored glove worn by the human teammate over time. We use classification (left cup vs. right cup intent) based on thresholds on the motion vector direction where the thresholds were tuned empirically for a table top interaction, face to face with the robot.

IV. MOTION PLANNING

Following the formalism defined by Dragan in [12], two motion planners are used in this study: a predictable planner and a legible planner. Predictable motion is the motion that is most efficient according to some cost function C over the trajectory ξ and can be generated through trajectory optimization. Given that our experimental task will be performed in a largely obstacle free environment, all predictable motions are set to be straight line trajectories from the start to goal.

Legible motion, in contrast, should aid the observer in the inference of the intended goal (action-to-goal inference). This requires modeling the observer's probability distribution over goals given trajectory segments. Following [12], a trajectory's legibility score is the normalized weighted summation of probabilities assigned to the robot's goal, G_R, across the trajectory with weights set according to $f(t)$:

$$\text{LEGIBILITY}[\xi] = \frac{\int P(G_R|\xi_{S\to\xi(t)})f(t)dt}{\int f(t)dt} \quad (1)$$

where $\xi_{S\to\xi(t)}$ denotes the trajectory segment from the starting configuration (S) to the configuration at time t ($\xi(t)$). In this work, we use $f(t) = T - t$ with T being the total time. This gives more weight to earlier parts of the trajectory. Thus, the optimal legible trajectory ξ^* can be generated through trajectory optimization such that

$$\xi^* = \arg\max_{\xi_{S\to G_R}} \text{LEGIBILITY}[\xi] \quad (2)$$

As shown in [12], this objective can be optimized through an iterative gradient ascent algorithm. We initialize the algorithm with a straight line path between the start and goal.

V. EXPERIMENTAL DESIGN

To investigate the effects of intent recognition and legible motion on human-robot collaboration, we propose a counterbalanced 2-way within-subjects study. We anticipate that both intent recognition and motion type will affect the collaboration along both objective and subjective measures. In addition, we expect users' perception of the robot's performance to be higher when both intent recognition and legible motion are present.

- *H1 - Objective Measures of Collaboration*:
  1) Legibility will improve objective measures of collaboration.
  2) Intent recognition will improve objective measures of collaboration.
- *H2 - Perceptions of Collaboration*:
  1) Legibility will positively affect perceptions of collaboration.
  2) Intent recognition will positively affect perceptions of collaboration.
- *H3 - Subjective Performance Rating*:
  1) Combined legible motion and intent recognition will be rated as better performing than either alone and over baseline.

A. Experimental Tasks

The task is setup as a pouring scenario where the robot and participant empty cups into the same container as shown in Figures 3 and 4 and the video attachment. We select this task because it is 1) collaborative, 2) requires both the robot and human recognize and communicate intent, and 3) repeatable across the conditions. For each round, the participant selects a cup that is either located on the right or left of the bins, and the robot tries to infer the correct side (intent recognition) in order to select the cup from the opposite side. The robot then empties its cup in one of the two bins first and the participant is required to empty his cup in the same bin as the robot's, inferring the correct bin via the robot's arm motion. Both the human and robot place the cups back in their original positions and then repeats the task. In order to enforce turn-taking, the participant is told that they must wait for the robot to take a photo of the scene. Upon hearing a camera shutter sound, the participant begins the next round by selecting a new cup. Each condition has four rounds. Neither the robot nor the participant know each other's goal *a priori*.

B. Independent Variables

The independent variables are intent recognition (absent vs. present) and motion type (predictable vs. legible). No intent recognition (IR Absent) means that when it is the robot's turn to select a cup, it makes its decision based on the available cups on the table, as per Section III. At this point, the participant would have already removed a cup from the table. The presence of intent recognition (IR Present) means that the robot predicts which cup the participant is going to grab before the participant actually grabs the cup, as per Section III. There were four conditions: Baseline (IR Absent and Predictable), IR Present and Predictable, IR Absent and Legible, IR Present and Legible.

C. Integrated Intent Recognition and Generation System

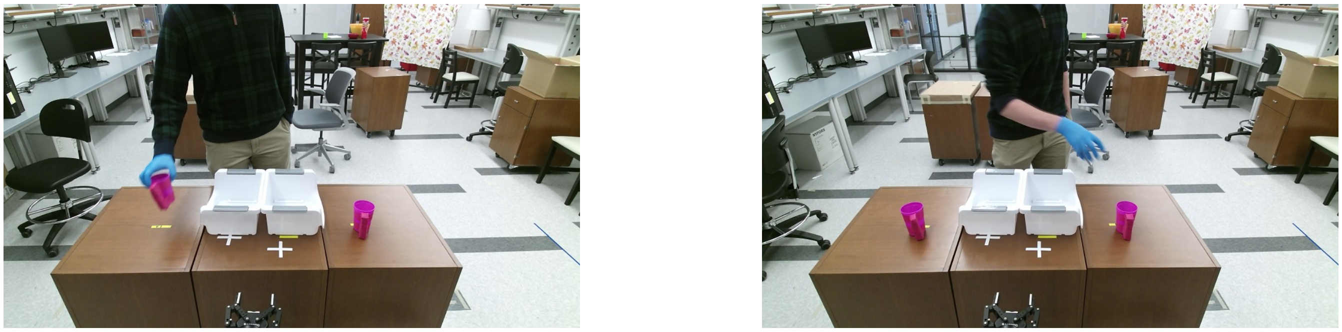
Testing the four conditions described in Section V-B requires an integrated intent recognition and generation system. The high-level system diagram is shown in Figure 2. This system is composed of four modules: segmentation, state extraction, intent recognition, and motion execution. The segmentation and state extraction modules are described in Section III.

The intent recognition module informs the motion execution module of the human's intent. Since intent recognition is one of our manipulated variables, this module is instantiated in one of two ways depending on the experimental condition: baseline object-based intent recognition (Section III) and hand motion-based intent recognition (Section III).

The motion execution module is triggered to pour from a cup upon intent recognition. Using the state information provided by the state extraction module, motion execution instructs the robot to grasp one of the remaining objects on the opposite side of the detected intent. The robot arm then returns to the home position from which it will then empty the cup in one of the two bins. The type of trajectory the robot executes is varied according to the experimental condition. All other trajectory segments display predictable motion. The robot's decision of which specific cup to grab and which bin to empty the cup in are randomly selected to prevent participants from guessing a pattern. All trajectories are pre-generated. The execution time for legible motion is 60.5 sec and predictable motion is 62.0 sec. This small difference of 2.5% is negligible compared to the total execution time.

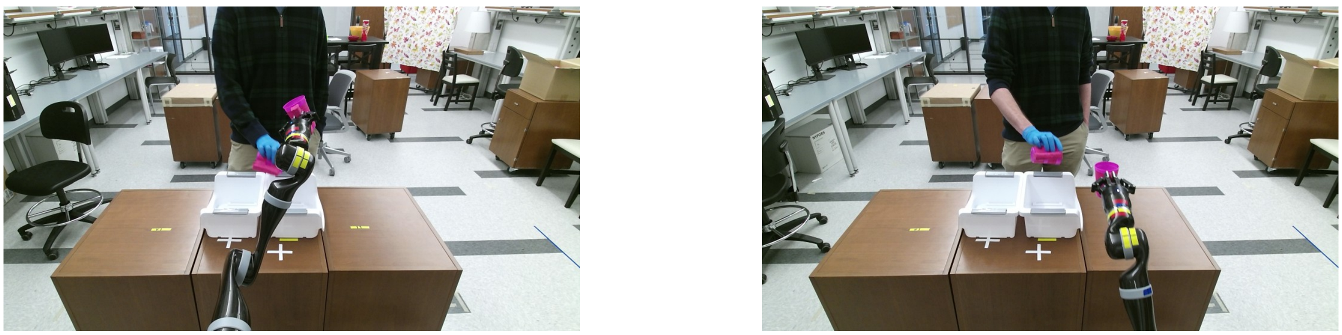
D. Participants

A total of 18 participants, 5 females and 13 males, participated in the study. All the participants were university students. To enable participants to compare the four conditions, the experiment used a within subjects design. The order of the conditioned were counterbalanced to control for order effects.



(a) Object Intent (b) Hand Motion Intent

Fig. 3. The two intent recognition modules detect intent at different times during the cup selection phase. The object intent system (a) detects intent only after the cup is grasped, while the hand motion intent system (b) can successfully detect intent before the cup is grasped.



(a) Predictable Motion (b) Legible Motion

Fig. 4. The two motion planners result in the human correctly identifying the robot's intended goal at different times during the bin selection phase. Under predictable motion (a), goal inference (bin selection) is delayed until the gripper is near the intended bin. Under legible motion (b), goal inference occurs earlier in the trajectory.

E. Procedure

First, participants were briefed on the collaborative task and were informed that four different robot "programs" were being tested. They were also asked to wear a blue glove during the duration of the study. The participants practiced the task once under the Baseline condition. Then, they performed the task under each condition. After each condition, they completed a brief questionnaire. At the end of the study, a post-study questionnaire was administered.

F. Dependent Measures

The dependent measures include both objective and subjective measures. The objective measures are:

- **Human's initial intent recognition time:** This is the amount of time it takes the human to initially infer the robot's bin selection. This time starts from the moment that the robot starts moving to the bin until the human starts moving their cup towards the predicted bin.
- **Human's final intent recognition time:** This is the time period where the human starts moving their cup towards the predicted bin and starts to pour the cup into the bin. The human's prediction of the robot's bin selection is confirmed when they start to pour the cup.
- **Percentage of overall concurrent motion:** This is the amount of concurrent motion divided by the total task time. The total task time is defined as the time when the human's hand starts moving to the start of the pouring which could be the human or the robot.
- **Percentage of segment concurrent motion:** This encompasses only the segment of the arm trajectory that have the predictable and legible components. This metric is calculated by dividing the total task time by the concurrent motion.

Most of the subjective measures are based on Dragan et al.'s [13] subset of questions from Hoffman's metrics for fluency in human-robot collaborations [14]. Table I shows the subjective scales that were used. Each of these with exception of the Post-Study question were rated on a 7-point Likert scale.

VI. RESULTS

A statistical model based on the 2 x 2 within-subjects design with motion type (MT) and intent recognition (IR) as factors was used in the analyses of variance (ANOVA). Post-hoc comparisons were conducted using Tukey HSD test. Data where the intent recognition system failed was excluded. Failure of the intent recognition system is defined as requiring the participant to reach more than twice to detect the hand motion.

We focus our analysis on the time period from the

Fluency	
1. The human-robot team worked fluently together.	
2. The robot contributed to the fluency of the team interaction.	
Robot Contribution	
1. I had to carry the weight to make the human-robot team better.	
2. The robot contributed equally to the team performance.	
Capability	
1. I am confident in the robot's ability to help me.	
2. The robot is intelligent.	
Legibility	
1. The robot can reason about how to make it easier for me to predict which bin it is reaching for.	
2. It was easy to predict the bin that the robot was reaching for.	
3. The robot moved in a manner that made its intention clear.	
4. The robot was trying to move in a way that helped me figure out which bin it was reaching for.	
Intent Recognition	
1. The robot can reason about what object I am reaching for	
2. I am confident that the robot can infer my intentions.	
The robot moved in a manner that made it clear it understood my intent.	
Post-Study	
1. Out of all the robot teammate programs, was there one that performed significantly better?	
2. If yes, please describe the program including how it performed significantly better.	

TABLE I

SUBJECTIVE MEASURES USED IN THE USER STUDY.

start of the task to the moment where the human starts pouring the cup. For the measures of human's initial intent recognition time, human's final intent recognition time, percentage of overall concurrent motion, and percentage of segment concurrent motion, two coders coded the video data. A high degree of reliability was found between them. The average measure Intra-Class Correlation (ICC) was 1 with a 95% confidence interval from 0.999 to 1 ($F(191,192) = 4433, p < 0.001$).

A. H1 - Objective Measures of Collaboration

Our analysis showed a marginally significant main effect of MT on the human's initial intent recognition time (MT: $F(1,17) = 3.183, p < 0.09$). For the human's final intent recognition time, there were no significant results. For the percentage of overall concurrent motion, the interaction was significant as shown in Figure 5 (MT by IR: $F(1,17) = 9.74, p < 0.01$). The effects of MT differs as a function of IR. The Legible Motion and IR Absent condition resulted in the most concurrent motion as compared to the Predictable and IR Absent condition. The interaction was marginally significant for the percentage of segment concurrent motion (MT by IR: $F(1,17) = 3.58, p = 0.07$); however, the post-hoc test did not yield any significant results.

B. H2 - Perceptions of Collaboration

The scales shown in Table I were combined and analyzed with ANOVAs and the results are in Figure 6. Participants ratings of team fluency was influenced by MT (MT: $F(1,17) = 7.72, p < 0.05$). Participants thought team fluency was higher when the robot displayed legible motion.

For ratings of robot contribution, only the main effect of MT was significant ($F(1,17) = 6.44, p < 0.05$). Similarly,

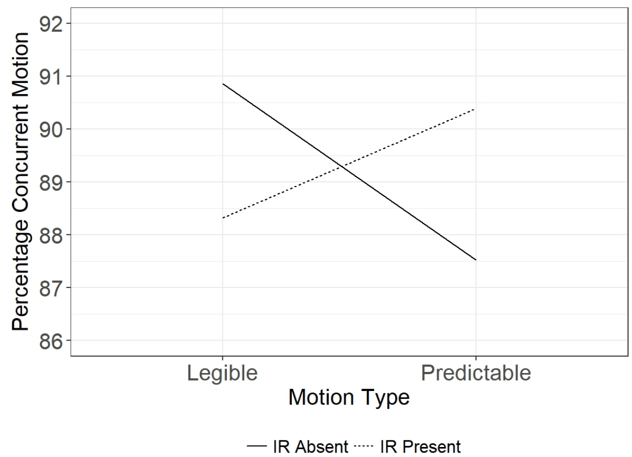


Fig. 5. The interaction effect between motion type and intent recognition was significant. The Legible Motion and IR Absent condition resulted in the most concurrent motion as compared to the Predictable and IR Absent condition.

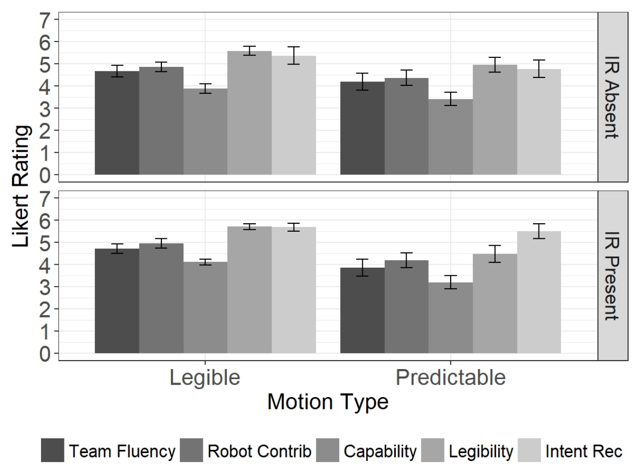


Fig. 6. The overall results of the subjective scales show that participants gave higher ratings when the robot displayed legible motion.

Figure 6 shows that for ratings of robot capability, the omnibus ANOVA indicated a significant main effect of MT, $F(1,17) = 12.36, p < 0.01$. Participants perceived the robot's contribution and capability to be higher with legible motion. As expected, MT significantly affected the legibility rating, $F(1,17) = 8.79p < 0.01$. The omnibus ANOVA also indicated a significant interaction between MT and IR, $F(1,17) = 4.59, p < 0.05$.However, results from the post-hoc analysis were not significant. Furthermore, the omnibus ANOVA indicated a marginally significant main effect of IR for the intent recognition rating, $F(1,17) = 3.84, p = 0.06$.

C. H3 - Subjective Performance Rating

Results from the post-study questionnaire showed that 14 out of 18 participants (78%) thought one of the four robot programs performed significantly better. In terms of which program performed significantly better, the voting results are: 50% voted for legible motion and intent recognition, 28% for legible motion and no intent recognition, 14% for predictable motion and intent recognition and 7% for the baseline.

VII. DISCUSSION

In this work, we presented results from a user study that investigated the effects of both intent recognition and communication in a human-robot collaborative cup pouring task. We showed that legible motion results in significant improvements in most of the subjective human-robot team performance measurements which supports hypothesis H1-1. Legible motion positively influenced participants' perception of team fluency, robot contribution, robot capability, and legibility, consistent with the results of prior work [13].

Qualitatively, we observed that rather than trying to finish the task quickly, many participants attempted to time their pouring action to happen concurrently with the robot's, likely as an attempt at collaboration. Thus, we analyzed the amount of human-robot concurrent motion as a way to capture any difference in the team's ability to do this coordination across conditions. For overall concurrent motion, we found a significant interaction effect, with the baseline condition having the least concurrent motion. That is, participants received cues about the task later, and had to wait longer before they could start moving to synchronize with the robot. The most concurrent motion occurred with predictable motion when IR was present, and legible motion when IR was absent. That is, the most coordination occurred when the robot started moving earlier and moved quickly to the goal, or when the robot waited to move and moved indirectly to the goal. This may be because in those conditions, the robot's behavior encouraged consistency in speed (or slowness), and the participants attempted to join in that consistency. The significant interaction effect for overall concurrent motion suggests that intent recognition and communication should be studied as an integrated system.

The use of intent recognition did not result in significant improvements in the performance measurements. Participants did notice a difference in the two intent recognition system as reported in the survey, where the majority thought the condition with legible motion and intent recognition performed significantly better, thus supporting H3. These results suggest an integrated intent recognition and communication system may promote more collaborative behavior among team members.

In this work, we experimented with a simple collaborative pouring task. A look at more complex tasks such as those involving multi-tasking and time pressure may be interesting to see if the same results would hold.

These results, especially the significance of the interaction effects, highlight the importance of considering coordination behaviors such as legible motion and intent recognition in combination as well as independently. Additionally, the results of our experiment show that people are drawn to collaboration and synchronization, and that they may not always be optimizing for speed in collaborative interactions. A promising direction for future research is to study these interactions between collaboration, timing, and robot motion.

VIII. CONCLUSION

This work is a first step towards exploring the effects of integrated intent recognition and legible motion on human-robot collaboration. Our initial findings suggest that a robot that can both recognize and communicate intent is more likely to increase collaboration in the team and thus enhance the team performance. Legible motion also positively influenced participants' perception of the robot and team.

ACKNOWLEDGMENT

The authors thank Xing Han and Kai Chih Chang for their contributions to the initial idea of this work.

REFERENCES

[1] J. R. Hoare and L. E. Parker, "Using on-line conditional random fields to determine human intent for peer-to-peer human robot teaming," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. IEEE, 2010, pp. 4914–4921.

[2] H. S. Koppula and A. Saxena, "Anticipating human activities using object affordances for reactive robotic response," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 1, pp. 14–29, 2016.

[3] J. Mainprice and D. Berenson, "Human-robot collaborative manipulation planning using early prediction of human motion," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*. IEEE, 2013, pp. 299–306.

[4] K. P. Hawkins, S. Bansal, N. N. Vo, and A. F. Bobick, "Anticipating human actions for collaboration in the presence of task and sensor uncertainty," in *Robotics and automation (ICRA), 2014 ieee international conference on*. IEEE, 2014, pp. 2215–2222.

[5] C. Pérez-D'Arpino and J. A. Shah, "Fast target prediction of human reaching motion for cooperative human-robot manipulation tasks using time series classification," in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*. IEEE, 2015, pp. 6175–6182.

[6] C.-M. Huang and B. Mutlu, "Anticipatory robot control for efficient human-robot collaboration," in *Human-Robot Interaction (HRI), 2016 11th ACM/IEEE International Conference on*. IEEE, 2016, pp. 83–90.

[7] J. Lasseter, "Principles of traditional animation applied to 3d computer animation," in *ACM Siggraph Computer Graphics*, vol. 21, no. 4. ACM, 1987, pp. 35–44.

[8] M. J. Gielniak and A. L. Thomaz, "Enhancing interaction through exaggerated motion synthesis," in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. ACM, 2012, pp. 375–382.

[9] ——, "Spatiotemporal correspondence as a metric for human-like robot motion," in *Proceedings of the 6th international conference on Human-robot interaction*. ACM, 2011, pp. 77–84.

[10] M. J. Gielniak, C. K. Liu, and A. L. Thomaz, "Generating human-like motion for robots," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1275–1301, 2013.

[11] D. Szafir, B. Mutlu, and T. Fong, "Communicating directionality in flying robots," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2015, pp. 19–26.

[12] A. Dragan and S. Srinivasa, "Integrating human observer inferences into robot motion planning," *Autonomous Robots*, vol. 37, no. 4, pp. 351–368, Dec 2014. [Online]. Available: https://doi.org/10.1007/s10514-014-9408-x

[13] A. Dragan, S. Bauman, J. Forlizzi, and S. Srinivasa, "Effects of robot motion on human-robot collaboration," in *Human-Robot Interaction*, Pittsburgh, PA, March 2015.

[14] G. Hoffman, "Evaluating fluency in human-robot collaboration," in *International conference on human-robot interaction (HRI), workshop on human robot collaboration*, vol. 381, 2013, pp. 1–8.