Task-Oriented Robot-to-Human Handovers in Collaborative Tool-Use Tasks

Meiying Qin[1] and Jake Brawer[1] and Brian Scassellati[1]

***Abstract*— Robot-to-Human handovers are common exercises in many robotics application domains. The requirements of handovers may vary across these different domains. In this paper, we first devised a taxonomy to organize the diverse and sometimes contradictory requirements. Among these, task-oriented handovers were not well-studied but important because the purpose of the handovers in human–robot collaboration (HRC) is not merely to pass an object from a robot to a human receiver, but to enable the human receiver to use it in a subsequent tool-use task. A successful task-oriented handover should incorporate task-related information – orienting the tool such that the human can grasp it in a way that is suitable for the task. We identified multiple difficulty levels of task-oriented handovers, and implemented a system to generate task-oriented handovers with novel tools on a physical robot. Unlike previous studies on task-oriented handovers, we trained the robot with tool-use demonstrations rather than handover demonstrations, since task-oriented handovers are dependent on the tool usages in the subsequent task. We demonstrated that our method can adapt to all difficulty levels of task-oriented handovers, including tasks that matched the typical usage of the tool (level I), tasks that required an improvised and unusual usage of the tool (level II), and tasks where the handover was adapted to the pose of a manipulandum (level III). We evaluated the generated handovers with online surveys. Participants rated our handovers to appear more comfortable for the human receiver and more appropriate for subsequent tasks when compared with typical handovers from prior work.**

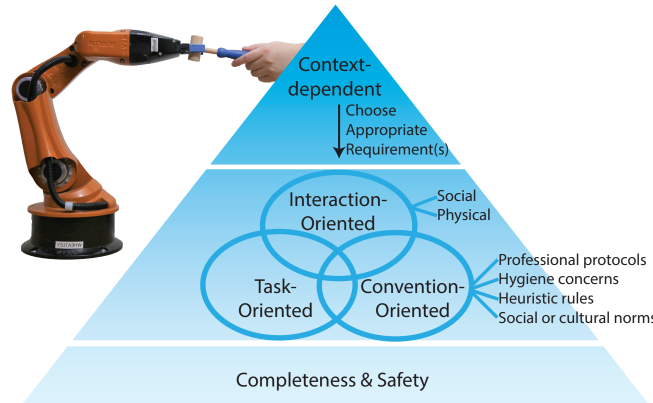

Fig. 1: Our taxonomy of robot-to-human handover requirements. Bottom to top: the basic, intermediate and advanced requirements.

I. INTRODUCTION AND RELATED WORKS

A robot-to-human handover is a joint action wherein a robot grasps, presents, and transfers an object held in its end-effector to a human receiver. It is a common exercise in numerous applications, including service robots handing flyers to pedestrians [1], personal assistive robots handing phones to people with disabilities [2], and factory robots handing hammers to collaborators [3]. To summarize the different requirements for handovers, we compiled a robot-to-humman handover taxonomy (for details, see Section I-A). The taxonomy serves the following purposes: 1) it helps to situate our study in the larger picture of robot-to-human handovers; 2) it helps to organize related work on handovers; 3) it may serve as a guide for future systems designed for handovers in terms of what requirements may need to be considered.

This study focused on one specific handover, the *task-oriented handover* that is commonly seen in the context of human-robot collaboration (HRC). However, as mentioned in recent publications [4], [5], task-oriented handovers have not yet gained enough attention in robot manipulations. In HRC, the purpose of a task-oriented handover typically is not merely to pass an object to a human, but also to enable the human to use the object to complete tasks. In order to maximize efficiency, the task-oriented handover should allow the human receiver to initiate a subsequent task with minimum in-hand object adjustment. Consequently, handovers of this type are dependent on how the tools should be used. Previous studies on task-oriented handovers generally demonstrated handovers of certain tools, without providing information regarding how the tools are used in the subsequent tasks. As a result, robots' lack of understanding of tool-use impedes their ability to generate handovers with novel tools. Therefore, our study aimed at designing a system that can generate appropriate task-oriented handovers with demonstrations of tool-use rather than handovers by integrating existing techniques. Furthermore, we also identified multiple levels of difficulties in task-oriented handovers and organized related work accordingly (for details, see Section I-B).

We built a system that generates task-oriented handovers. The system learned tool-affordances to allow the robot to understand the nature of the subsequent task. In our system, we chose and integrated a tool-affordance learning technique appropriate for handover tasks. We implemented the system on a physical robot and the results showed that the system can handle all difficulty levels of task-oriented handovers. We also conducted an online survey to evaluate the handovers executed by the robot. In summary, our contributions are:

[1]The authors are with Department of Computer Science, Yale University, New Haven, CT, United States meiying.qin@yale.edu, or meiyingqin@gmail.com

1) We defined a taxonomy of handover requirements.
2) Our system generated handovers based on learned tool affordances, rather than handover demonstrations, since task-oriented handovers are dependent on the subsequent tool-use task.
3) With the understanding of how tools should be used, our system was able to handle task-oriented handovers for all three difficulty levels that we identified.
4) Survey participants preferred our handovers and rated them as appearing to be comfortable for a potential human receiver and appropriate for the subsequent tasks.

A. Taxonomy: Handover Requirements

A handover is a complex manipulation with various requirements to satisfy. Therefore, we compiled a taxonomy of handover requirements and summarized it in Fig. 1. The requirements at the lower levels should be satisfied first before a higher-level requirement can be satisfied. In the taxonomy, the basic requirement is to be complete and safe. A complete handover refers to the successful delivery of an object to the receiver [6], [7], [8], [9], [10], [11], [12], [13], and a safe handover requires that no collision occurs at any time during the course of delivery [14], [15], [16], [17]. This is the focus of most handover studies.

Beyond the basic requirement of completeness and safety, satisfying one or more intermediate requirements will produce appropriate handovers. Compared with the studies focused on basic requirements, fewer handover studies focus on intermediate requirements.

The first intermediate requirement is that handovers should adapt to social or physical interactions between a human receiver and a robot (i.e., interaction-oriented). The social interactions include sending or perceiving various types of social signals such as eye contact [18], [19], [20], [21], while the physical interactions involve adjusting where [22], [23] or when [24] to conduct handovers based on the location or the physical state (e.g., availability) of a human receiver, or generating handovers that comply with human ergonomic needs [25], [26]. Satisfying these interaction–oriented requirements can help with generating customized handovers that are more comfortable for the receiver.

The second intermediate requirement is that a handover should abide by various conventions (i.e., convention-oriented), including professional protocols (e.g., handing over a surgical tool to a surgeon during a procedure in the operating room), hygiene concerns (e.g., one should not grasp the tines of a fork which will touch food), heuristic rules (e.g., one tend to orient an object horizontally for the receiver), and social or cultural norms (e.g., handing over a gift with a single hand is considered disrespectful). Satisfying convention-oriented requirements can help with generating handovers that match expectations.

The third intermediate requirement is that handovers should incorporate information about subsequent tasks (i.e., task-oriented) [27], [28], [5], [29], which allows the human receiver to perform the subsequent tasks more efficiently.

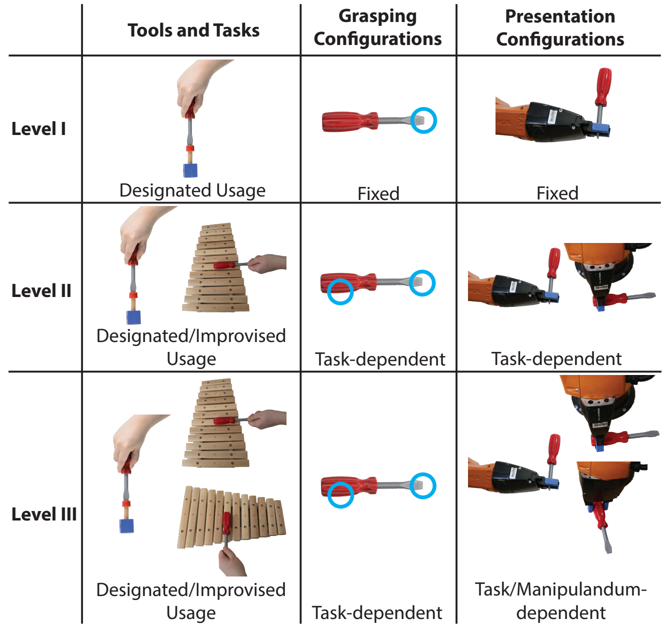

Fig. 2: Difficulty levels of task-oriented handover. Blue circles indicate function parts for different tasks.

Our study focus on this third intermediate requirement, task-oriented handovers, and other requirements are beyond the scope of this study.

The advanced requirement in our taxonomy is that a handover should be context-dependent. In other words, one should choose one or a combination of intermediate requirements to meet based on the specific context. The intermediate requirements may contradict each other, and not all requirements can be satisfied simultaneously. For example, during a convocation, an assistant hands the diploma to a dean in a way that prioritizes the interaction-oriented requirements so that the dean can receive the diploma more comfortably. However, when the dean hands the diploma to a graduate, the dean will not prioritize the interaction-oriented requirements as the assistant does, but will prioritize the convention-oriented requirements and use both hands to show respect. Therefore, a robot needs to recognize which intermediate requirements are important in the given context and choose one or a combination of intermediate requirements to meet the given context.

B. Task-oriented Handovers

As the objects to be handed over in task-oriented handovers are usually *tools*, we consider task-oriented handovers in the context of tool-use, and the object manipulated by a tool is referred to as *manipulandum* in this paper.

We identified three levels of difficulties in task-oriented handovers and organized related work on task-oriented handovers accordingly. Fig. 2 summarizes the difficulty levels and shows examples of each level. Level I is to properly hand over a tool to a human to perform a task typically matched with the tool (e.g., using a screwdriver to drive screws). Since a tool usually has a default usage, level I handovers could be achieved by building or learning a dataset to store handovers [27], [28], [5], and the dataset was learned with handover demonstrations rather than tool-use demonstrations.

In level II task-oriented handovers, a human receiver may use tools with their default usages, but may also improvise tool-use for tasks not generally associated with the tools (e.g., using a screwdriver to play a xylophone rather than to drive a screw). It is more challenging than level I because a pre-built dataset that can handle level I handovers may not be able to handle level II handovers due to the nearly limitless ways any particular tool can be used in different tasks. More importantly, the dataset may not be able to generalize to level II handovers due to a lack of understanding of how the tools should be used. To realize handovers at this level, a robot should recognize the functional segment of the tool and understand the usage to determine the handovers. In other words, learning *tool affordance* is the key to achieving level II task-oriented handover. To our knowledge, only one previous study considered learning tool affordances before performing handovers [29]. Although only level I handovers were demonstrated, their system may be capable of level II handovers. However, the design of this previous study makes their system impractical to be applied in many HRC scenarios. In this previous study, a human needed to demonstrate the usage of the novel tool to the robot in order to determine relevant handover configurations. However, a novel tool to be handed over is generally out of reach of the human receiver, so that a demonstration may be impossible without handing over the tool in the first place.

In addition to level II handover constraints, a robot should adjust the handover configurations based on the pose of the manipulandum (i.e., level III handovers). While some tasks impose consistent orientations irrespective of the tool used (e.g., stirring a pot of broth always requires a vertical tool orientation), the usages of tools in other tasks depend on the pose of the manipulandum (e.g., using a screwdriver to drive a screw placed either vertically into a tabletop or horizontally into a wall). This imposes challenges for previous systems [29] because each task was bound with specific handover configurations. Therefore, tool affordance may need to be learned in a different way to allow level III task-oriented handovers.

Previous studies on tool affordance have learned tool-use in various ways. However, they may not be appropriate for task-oriented handovers. Tool affordances were learned as a distribution of the outcomes [30], [31] instead of the relationship between a movement of a tool and the corresponding status change of a manipulandum. With tool affordances learned in this manner, a robot cannot achieve level III handovers because the relation between specific usages and specific contexts is unknown. When the abovementioned relationship was learned in a previous study [32], it learned in a way that was specific to the learned tools, and it was unknown whether a robot could generalize the learned tools to novel tools. It would be tedious to learn to use every tool prior to handing it over. While parallel Self-Organizing Maps (SOMs) can help with handling novel tools, novel tools needed to share similar shapes with the training tools [33], imposing restrictions on what kinds of novel tools a robot could hand over. This problem was overcome by using a large training set [34], [35], [36], which may be impractical in time-sensitive scenarios to hand over tools.

II. DESIGN AND IMPLEMENTATION

In our system, a robot first learned tool affordances or how to use a tool. Then in a robot-to-human handover task, the handovers were calculated based on how a tool should be used in subsequent tasks, and were then passed on to standard inverse kinematics and motion planning libraries to execute the motion. The tools may even be novel such that the robot never observed their usages in the required task. In this case, the robot first inferred its usage based on how the tools were used in the same task, and then generated corresponding handovers.

A. Object Model Generation

Preliminary 3D models of the objects were scanned by the robot if possible. A script that utilized MeshLab[1] was used to automatically process the 3D models to smooth, upsample, recenter, and resurface the point clouds into triangular meshes. The 3D models of the tools were then segmented using the shape diameter function (CDF). The objects that could not be scanned by the robot were obtained manually with Autodesk Recap Pro[2]. Detailed procedures for obtaining 3D models can be found in our previous work [37].

B. Vision Module

To obtain the pose of an object in the scene, a partial point cloud of the object needs to be extracted from the environment. To isolate the partial point cloud, a background point cloud without the object and a foreground point cloud with the object was captured from a depth sensor. Both point clouds were processed to leave only the workspace, and the desktop was removed with random sample consensus (RANSAC). The partial point cloud of the object was obtained by subtracting the processed background point cloud from the processed foreground point cloud.

After obtaining the partial point cloud of the object, the pose of the object was retrieved by registering the partial point cloud with the triangular mesh of the object using a modified Iterative Closest Point registration (ICP) algorithm. In this study, the pose of an object was represented with a 4×4 homogeneous transformation matrix $T \in SE(3)$ (superscript: reference frame, subscript: object), and $SE(3)$ represents the special Euclidean group:

$$T = \begin{bmatrix} R & p \\ 0 & 1 \end{bmatrix}$$

where R is a 3×3 rotation matrix representing the orientation, and p is a vector representing the position. The pose of the tool $T^{world}_{tool\_on\_desk}$ and the manipulandum $T^{world}_{manipulandum}$ in the world frame were perceived when they were placed on the desktop.

[1]MeshLab: https://www.meshlab.net/

[2]Autodesk software: https://www.autodesk.com/

C. Learning Tool Affordances

The system learned tool affordances with the tool-use framework called TRansferrIng Skilled Tool-use Acquired Rapidly (TRI-STAR) framework [37]. While other systems can be used in place of TRI-STAR, we chose this system because it is flexible with the form-factors of the tools, can accommodate a wide range of tasks, is data-efficient, and is able to generalize to novel tools and manipulanda without extra training. Detailed methods can be found in our previous work [37].

The tool-use framework includes a tool-use task taxonomy based on the goal state of the manipulandum in different frames of reference. It learns how a tool acts upon a manipulandum in a task using Learning from Demonstration (LfD). Based on the demonstrations, the framework classifies the tasks according to the task taxonomy and learns the tool-use skills or tool affordances accordingly.

In TRI-STAR, the tool affordances include motor skills and contact poses. Motor skills include kinematics skills, such as a trajectory that a tool should follow, and dynamics skills which considers the forces. Though TRI-STAR currently only considers kinematics skills, dynamics skills are less relevant in handover tasks as the robot may not need to know the force that needs to be exerted while the human collaborator using the tool in order to find the appropriate handover configurations. The other component of tool affordances, which is the contact pose, include the grasping pose of the tool and the tool-manipulandum contact poses while using the tool. The grasping pose is dependent on the tool-manipulandum contact poses. Each segment between the demonstrated key points of the trajectories is represented with exponential representations that parametrize the segment with a screw axis and an angle. The segments are then grouped based on similar screw axes. As a result, the entire trajectory is represented with a series of pairs of a screw axis and an angle. The contact pose is represented by a class. Poses in the same class can be obtained by rotating about an axis. Based on the demonstrations, the framework needs to calculate the axis, choose one pose as the starting pose, and decide the range of rotation allowed about the axis. The range of the rotation depends on the type of task. For example, a slotted screwdriver may contact a slotted screw in two ways, while a hammer may approach a nail from infinitely many directions. Though the representations of the kinematics skills and contact skills are relatively uniform across all tasks, the choice of the frames of reference is dependent on the type of tasks in the task taxonomy.

Given novel tools and manipulanda, the key is to find how the object should substitute the learned object. In other words, the system should find the pose of the novel object in the reference frame of the learned object when using the objects in the tool-use task. The substitution is calculated by aligning the source objects and novel objects based on the global or local geometric features. When aligning the objects for global features, the point cloud is stretched or compressed disproportionally along different axes so that the bounding boxes of the objects match. The point cloud of the source object and the substitute object is mapped via modified ICP in order to gain the best matching result. When aligning the objects for local features, the functional part of the object is stretched or compressed proportionally so that the longest edges of the bounding boxes match. The functional part of the source and substitute object is then mapped via modified ICP. In this way, two substitutions are obtained. One optimizes the global shape, and one optimizes the local feature. The system chooses the substitution with a better matching result from these two options.

D. Grasping Configurations

The grasping configuration, which is the end-effector pose $T^{world}_{ee\_grasp}$ when grasping the tool, includes the orientation $R^{world}_{ee\_grasp}$ and the position $p^{world}_{ee\_grasp}$ of the end-effector.

1) Grasping Orientations: The tool to be handed over was assumed to be resting on the desk for simplicity. The grippers grasped the tool from above with the fingers perpendicular to the desktop. The opening of the gripper should be perpendicular to the primary axis $\vec{pa}$ of the tool (i.e., the direction of the longest edge of the minimum bounding box of the object), which resulted in the orientation $R^{tool\_adjusted}_{ee\_grasp}$ of the gripper being unchanged in the adjusted tool frame. Given the perceived pose of the tool $T^{world}_{tool\_on\_desk}$, the x axis of the adjusted tool frame $R^{world}_{tool\_adjusted}$ was defined as the unit primary axis of the tool, the z axis was defined as the unit vector opposite to the direction of standard gravity, and the y axis was calculated using the right-hand rule. With the adjusted tool frame, the orientation of the end-effector $R^{world}_{ee\_grasp}$ was calculated as (where $\times$ is matrix multiplication):

$$R^{world}_{ee\_grasp} = R^{world}_{tool\_adjusted} \times R^{tool\_adjusted}_{ee\_grasp}$$

2) Grasping Positions: The grasping position of the end-effector $p^{world}_{ee\_grasp}$ was initially chosen as the center of the contact area $p^{world}_{tool\_contact}$ of the tool when used on a manipulandum, because the contact area was the part of the tool least likely to be the handle. With learned tool affordances, the TRI-STAR framework calculated the contact area of the tool based on the manipulandum and the subsequent task. The center of the contact area $p^{model}_{tool\_contact}$ was calculated as the center of the minimum bounding box of the contact area. The use of a bounding box reduced bias due to the density of a point cloud. The $p^{world}_{tool\_contact}$ was obtained using $T^{world}_{tool\_on\_desk} \times p^{model}_{tool\_contact}$. To ensure stable grasping, the fingers of the grippers should distribute evenly around the primary axis of the tool. The grasping position needed to be adjusted by projecting $p^{world}_{tool\_contact}$ onto the primary axis $\vec{pa}$ to obtain an adjusted grasping position $p^{world}_{tool\_adjusted\_contact}$:

$$\begin{aligned} p^{world}_{tool\_adjusted\_contact} = & \frac{(p^{world}_{tool\_contact} - p^{world}_{tool\_center}) \cdot \vec{pa}}{\|\vec{pa}\|^2} \vec{pa} \\ & + p^{world}_{tool\_center} \end{aligned}$$

where $p^{world}_{tool\_center}$ was the center of the minimum bounding box of the tool. The grasping position $p^{world}_{ee\_grasp}$ was set to be

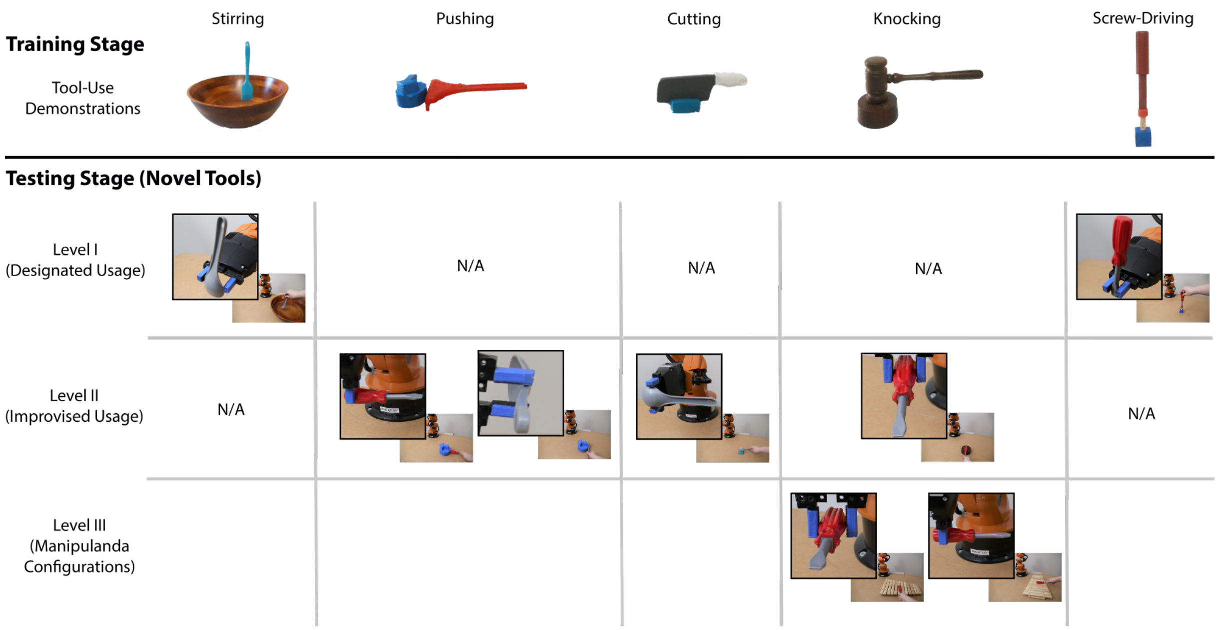

Fig. 3: Handover evaluations on five tool-use tasks. (Top) The system was first trained with how to perform the stirring, pushing, cutting, knocking, and screw-driving tasks, rather than demonstrations of handovers. (Bottom) The robot was required to generate handovers for the human receiver to perform subsequent tasks. The handovers were with different levels of difficulty. The 'N/A' either refers to that the tool cannot perform the handover at the difficulty level, or the tool is inappropriate. Each cell shows a demonstration, which shows the handover generated and how the human used the tool to perform the subsequent task. The pictures were taken from the view of the human receiver.

$p^{world}_{tool\_adjusted\_contact}$ and the z was set to be the value where the gripper just touched the desktop.

E. Presentation Configurations

Presentation configurations are the end-effector poses when the robot presents the tool to the human collaborator to grasp it. In order to minimize in-hand tool adjustment, the orientation of the tool $R^{world}_{tool\_present}$ should be close to the orientation when the human receiver started to use the tool $R^{world}_{tool\_usage}$, while the location of the handover $p^{world}_{tool\_present}$ was pre-set since the human receiver was assumed to be at a fixed location. Each $T^{world}_{tool\_present}$ corresponding to a $T^{world}_{ee\_candidate}$ was calculated using $T^{world}_{ee\_candidate} \times T^{ee}_{tool}$ where $T^{ee}_{tool} = {T^{world}_{ee\_grasp}}^{-1} \times T^{world}_{tool\_on\_desk}$ since the tool was grasped securely so that T^{ee}_{tool} was unchanged. $R^{world}_{tool\_usage}$ was selected to be the start orientation of the tool trajectory in the world frame. The start orientation $R^{manipulandum}_{tool\_usage}$ of the tool trajectory in the manipulandum frame was generated from the TRI-STAR framework with learned tool affordances. $R^{world}_{tool\_usage}$ was obtained using $R^{world}_{manipulandum} \times R^{manipulandum}_{tool\_usage}$.

III. Experiments

We implemented and tested our system on a Kuka youBot robot without the mobile base. A Microsoft Azure RGB-D camera placed on the side of the workspace was used to perceive the pose of the tools and manipulanda. The human receiver was assumed to stand at a fixed location since adapting to the human location is beyond the scope of this study. In the training stage, the robot was trained with twenty demonstraions per tool in simulation to learn the tool affordances or how to use these tools in five tasks (i.e., stirring, pushing, cutting, knocking, and driving screws). The training tools and manipulanda are shown in Fig. 3. No additional training was needed to perform handovers after learning the tool affordances. In the testing stage, the robot was required to hand over novel tools to a human to complete tasks and it was informed which task that the human receiver would perform.

A. Robot Validations

We conducted two experiments. Experiment I tested how the robot handed over a novel tool to complete tasks required either typical (i.e., level I) or improvised (i.e., level II) usage of the tool. Experiment II tested how the robot handed over a novel tool to complete a task with different manipulandum poses (i.e., level III). In order to show that the system can generate different handover configurations of a tool for different subsequent tasks, we chose the same tool to perform as many tasks as possible in the testing phase rather than one novel tool in each task. In experiment I, a spoon and a screwdriver were chosen as the novel tools. As shown in Fig. 3, the human receiver was required to perform the stirring, pushing, and cutting tasks with the spoon, and to perform the pushing, knocking, and driving screws tasks with the screwdriver. A single tool was not required to perform all tasks because some tasks were inappropriate for the tool.

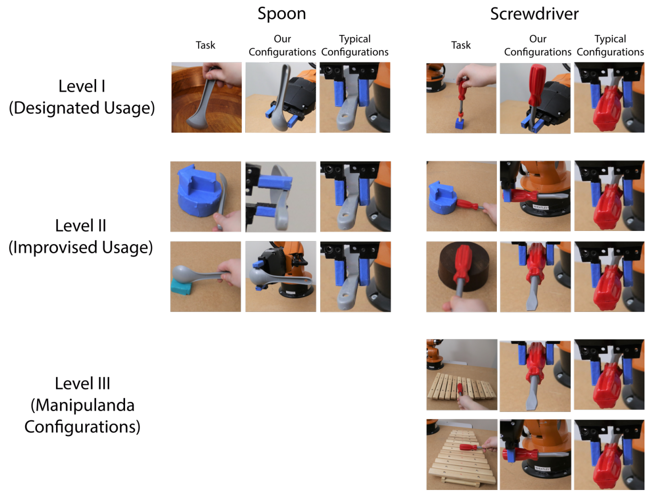

Fig. 4: Comparing handover configurations generated by our system and the typical handovers in previous studies. The figure includes handovers of level I (top), level II (middle), and level III (bottom), with the spoon (left) and with the screwdriver (right) in different tasks. The typical handovers always grasp the same location on a tool and orient the handle of a tool horizontally to the human receiver. In contrast, our configurations are customized to the subsequent tasks and thus require minimum in-hand tool adjustments for the human receiver.

In experiment II, even the manipulanda is a novel object, a xylophone, while the tool is the screwdriver. The xylophone was placed with two different orientations.

The results showed that the robot was able to handle level I, level II and level III handovers by adjusting both the grasping and presentation configurations according to the tasks. We compared our configurations with the typical configurations in previous studies as shown in Fig. 4. While our handover configurations were customized to the subsequent tasks, typical configurations in previous studies followed heuristic rules that a robot always selected a fixed location on the tool to grasp and oriented the handle horizontally towards the human receiver to present it. Therefore, handovers using our configurations required minimum in-hand tool adjustments when compared with the handovers using typical configurations.

B. Survey

To evaluate how naïve end-users perceive handovers generated by the robot, we conducted a survey on Amazon Mechanical Turk. Informed consent was obtained electronically. We recruited 70 participants, and each was compensated with \$5. Out of the 70 participants, 15 were excluded from data analysis due to failing sanity-check questions. The data from the 55 eligible participants (35 males, 20 females) with an average age of 35.6 years were analyzed. In the survey, the questions were randomized, as were the options in each question. We designed multiple-choice questions (MCQ) and rating questions, which showed pictures or videos of the handovers and how the human receiver uses the tool in the subsequent tasks. The pictures and videos were taken from the view of the human receiver. A sample of the questionnaire can be found here[3]. The MCQ responses were converted to continuous variables and were analyzed with one-sample t-tests to compare with the chance level. The ratings were analyzed with paired samples t-tests.

For experiment I, the participants chose our handovers over the handovers in previous studies 88% of the time ($t(54) = 13.843$, $p < .001$). They were able to predict the subsequent tasks correctly 79% of the time ($t(54) = 11.461$, $p < .001$) given our handovers. On a five-point Likert scale, participants rated our configurations (M = 4.38, SD = 0.87) being more appropriate ($t(54) = 5.650$, $p < .001$) for the subsequent task than the typical configurations (M = 3.04, SD = 1.23). They also rated our handovers (M = 4.34, SD = 0.91) to be more comfortable ($t(54) = 5.751$, $p < .001$) for the human receiver than the handovers in previous studies (M = 3.22, SD = 1.13), and the collaboration was perceived to be more fluent ($t(54) = 4.810$, $p < .001$) when the robot used our handovers (M = 4.31, SD = 0.86) than when using the typical configurations (M = 3.24, SD = 1.22). For experiment II, the participants chose preferred handover configurations from two options. Results showed that the participants preferred our handovers 82% of the time ($t(54) = 7.884$, $p < .001$).

IV. Conclusions

We compiled a taxonomy of different requirements for handovers in general, and identified three levels of difficulty for task-oriented handovers in particular. We also integrated a system for task-oriented handovers, and showed that the system was able to handle level I, level II, and level III task-oriented handovers, and thus made it possible for the human receiver to complete subsequent tasks more efficiently with diverse task specifications. Furthermore, the system was trained with tool affordances, rather than demonstrations of handovers, allowing the system to understand the tool-use tasks and generalize the handovers to novel tools. The online survey results showed that participants preferred our handovers over the typical handovers in previous studies.

Our system presents a contribution towards task-oriented handovers. However, we would like to acknowledge the limitations of the current study. We focused on task-oriented handovers, while other handover requirements are beyond the scope of this study. For example, this study focused on task-oriented handovers and did not consider other aspects such as adapting to social signals from the human. Moreover, we acknowledge that the conclusions based on online studies are limited compared with an in-person study.

ACKNOWLEDGMENT

This work is funded by the Office of Naval Research (ONR) award No. N00014-18-1-2776 and National Science Foundation (NSF) under grants No. 1955653, 1928448, 2106690, and 1813651.

[3]Survey: https://tinyurl.com/surveyforhandover

References

[1] C. Shi, M. Shiomi, C. Smith, T. Kanda, and H. Ishiguro, "A model of distributional handing interaction for a mobile robot," in *Proceedings of Robotics: Science and Systems*, Berlin, Germany, June 2013.

[2] Y. S. Choi, T. Chen, A. Jain, C. Anderson, J. D. Glass, and C. C. Kemp, "Hand it over or set it down: A user study of object delivery with an assistive mobile manipulator," in *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2009, pp. 736–743.

[3] A. Koene, S. Endo, A. Remazeilles, M. Prada, and A. M. Wing, "Experimental testing of the coglaboration prototype system for fluent human-robot object handover interactions," in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2014, pp. 249–254.

[4] V. Ortenzi, A. Cosgun, T. Pardi, W. P. Chan, E. Croft, and D. Kulić, "Object handovers: a review for robotics," *IEEE Transactions on Robotics*, 2021.

[5] V. Ortenzi, M. Controzzi, F. Cini, J. Leitner, M. Bianchi, M. A. Roa, and P. Corke, "Robotic manipulation and the role of the task in the metric of success," *Nature Machine Intelligence*, vol. 1, no. 8, pp. 340–346, 2019.

[6] M. Huber, M. Rickert, A. Knoll, T. Brandt, and S. Glasauer, "Human-robot interaction in handing-over tasks," in *RO-MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2008, pp. 107–112.

[7] A. Koene, A. Remazeilles, M. Prada, A. Garzo, M. Puerto, S. Endo, and A. M. Wing, "Relative importance of spatial and temporal precision for user satisfaction in human-robot object handover interactions," in *Third International Symposium on New Frontiers in Human-Robot Interaction*, 2014.

[8] W. P. Chan, C. A. Parker, H. M. Van der Loos, and E. A. Croft, "Grip forces and load forces in handovers: implications for designing human-robot handover controllers," in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, 2012, pp. 9–16.

[9] A. G. Eguiluz, I. Rañó, S. A. Coleman, and T. M. McGinnity, "Reliable object handover through tactile force sensing and effort control in the shadow robot hand," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 372–377.

[10] W. P. Chan, C. A. Parker, H. M. Van der Loos, and E. A. Croft, "A human-inspired object handover controller," *The International Journal of Robotics Research*, vol. 32, no. 8, pp. 971–983, 2013.

[11] J. Konstantinova, S. Krivic, A. Stilli, J. Piater, and K. Althoefer, "Autonomous object handover using wrist tactile information," in *Annual Conference Towards Autonomous Robotic Systems*. Springer, 2017, pp. 450–463.

[12] P. Neranon, "Robot-to-human object handover using a behavioural control strategy," in *2018 IEEE 5th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA)*. IEEE, 2018, pp. 1–6.

[13] S. Parastegari, E. Noohi, B. Abbasi, and M. Žefran, "A fail-safe object handover controller," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 2003–2008.

[14] M. Prada, A. Remazeilles, A. Koene, and S. Endo, "Implementation and experimental validation of dynamic movement primitives for object handover," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, pp. 2146–2153.

[15] G. J. Maeda, G. Neumann, M. Ewerton, R. Lioutikov, O. Kroemer, and J. Peters, "Probabilistic movement primitives for coordination of multiple human–robot collaborative tasks," *Autonomous Robots*, vol. 41, no. 3, pp. 593–612, 2017.

[16] E. A. Sisbot, L. F. Marin-Urias, X. Broquere, D. Sidobre, and R. Alami, "Synthesizing robot motions adapted to human presence," *International Journal of Social Robotics*, vol. 2, no. 3, pp. 329–343, 2010.

[17] J. Mainprice, E. A. Sisbot, T. Siméon, and R. Alami, "Planning safe and legible hand-over motions for human-robot interaction," in *IARP/IEEE-RAS/EURON workshop on technical challenges for dependable robots in human environments*, 2010.

[18] K. Strabala, M. K. Lee, A. Dragan, J. Forlizzi, S. S. Srinivasa, M. Cakmak, and V. Micelli, "Toward seamless human-robot handovers," *Journal of Human-Robot Interaction*, vol. 2, no. 1, pp. 112–132, 2013.

[19] H. Admoni, A. Dragan, S. S. Srinivasa, and B. Scassellati, "Deliberate delays during robot-to-human handovers improve compliance with gaze communication," in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, 2014, pp. 49–56.

[20] E. C. Grigore, K. Eder, A. G. Pipe, C. Melhuish, and U. Leonards, "Joint action understanding improves robot-to-human object handover," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 4622–4629.

[21] M. Cakmak, S. S. Srinivasa, M. K. Lee, S. Kiesler, and J. Forlizzi, "Using spatial and temporal contrast for fluent robot-human handovers," in *Proceedings of the 6th international conference on Human-robot interaction*, 2011, pp. 489–496.

[22] J. Aleotti, V. Micelli, and S. Caselli, "Comfortable robot to human object hand-over," in *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2012, pp. 771–776.

[23] K. L. Koay, E. A. Sisbot, D. S. Syrdal, M. L. Walters, K. Dautenhahn, and R. Alami, "Exploratory study of a robot approaching a person in the context of handing over an object." in *AAAI spring symposium: multidisciplinary collaboration for socially assistive robotics*. Stanford, CA, 2007, pp. 18–24.

[24] C.-M. Huang, M. Cakmak, and B. Mutlu, "Adaptive coordination strategies for human-robot handovers." in *Robotics: science and systems*, vol. 11. Rome, Italy, 2015.

[25] L. Peternel, W. Kim, J. Babič, and A. Ajoudani, "Towards ergonomic control of human-robot co-manipulation and handover," in *2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*. IEEE, 2017, pp. 55–60.

[26] A. M. Bestick, S. A. Burden, G. Willits, N. Naikal, S. S. Sastry, and R. Bajcsy, "Personalized kinematics for human-robot collaborative manipulation," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 1037–1044.

[27] W. P. Chan, M. K. Pan, E. A. Croft, and M. Inaba, "An affordance and distance minimization based method for computing object orientations for robot human handovers," *International Journal of Social Robotics*, vol. 12, no. 1, pp. 143–162, 2020.

[28] A. Bestick, R. Bajcsy, and A. D. Dragan, "Implicitly assisting humans to choose good grasps in robot to human handovers," in *International Symposium on Experimental Robotics*. Springer, 2016, pp. 341–354.

[29] W. P. Chan, Y. Kakiuchi, K. Okada, and M. Inaba, "Determining proper grasp configurations for handovers through observation of object movement patterns and inter-object interactions during usage," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, pp. 1355–1360.

[30] J. Sinapov and A. Stoytchev, "Detecting the functional similarities between tools using a hierarchical representation of outcomes," in *2008 7th IEEE International Conference on Development and Learning*. IEEE, 2008, pp. 91–96.

[31] A. Stoytchev, "Behavior-grounded representation of tool affordances," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2005, pp. 3071–3076.

[32] V. Tikhanoff, U. Pattacini, L. Natale, and G. Metta, "Exploring affordances and tool use on the icub," in *2013 13th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*. IEEE, 2013, pp. 130–137.

[33] T. Mar, V. Tikhanoff, G. Metta, and L. Natale, "Self-supervised learning of tool affordances from 3d tool representation through parallel som mapping," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 894–901.

[34] P. Gajewski, P. Ferreira, G. Bartels, C. Wang, F. Guerin, B. Indurkhya, M. Beetz, and B. Śniezyński, "Adapting everyday manipulation skills to varied scenarios," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 1345–1351.

[35] K. Fang, Y. Zhu, A. Garg, A. Kurenkov, V. Mehta, L. Fei-Fei, and S. Savarese, "Learning task-oriented grasping for tool manipulation from simulated self-supervision," *The International Journal of Robotics Research*, vol. 39, no. 2-3, pp. 202–216, 2020.

[36] A. Xie, F. Ebert, S. Levine, and C. Finn, "Improvisation through physical understanding: Using novel objects as tools with visual foresight," in *Proceedings of Robotics: Science and Systems*, FreiburgimBreisgau, Germany, June 2019.

[37] M. Qin, J. Brawer, and B. Scassellati, "Rapidly learning generalizable and robot-agnostic tool-use skills for a wide range of tasks," *Frontiers in Robotics and AI*, vol. 8, 2021.