Robot, Pass Me the Tool: Handle Visibility Facilitates Task-oriented Handovers

Valerio Ortenzi*
MPI for Intelligent Systems
Stuttgart, Germany
valerio.ortenzi@gmail.com

Maija Filipovica*
University of Bristol
Bristol, UK
qq20468@bristol.ac.uk

Diar Abdlkarim
University of Birmingham
Birmingham, UK
DXA392@student.bham.ac.uk

Tommaso Pardi
University of Birmingham
Birmingham, UK
txp754@student.bham.ac.uk

Chie Takahashi
University of Cambridge
Cambridge, UK
ct564@cam.ac.uk

Alan M. Wing
University of Birmingham
Birmingham, UK
a.m.wing@bham.ac.uk

Massimiliano Di Luca
University of Birmingham
Birmingham, UK
M.DiLuca@bham.ac.uk

Katherine J. Kuchenbecker
MPI for Intelligent Systems
Stuttgart, Germany
kjk@is.mpg.de

2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI) | 978-1-6654-0731-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/HRI53351.2022.9889546

***Abstract*—A human handing over an object modulates their grasp and movements to accommodate their partner's capabilities, which greatly increases the likelihood of a successful transfer. State-of-the-art robot behavior lacks this level of user understanding, resulting in interactions that force the human partner to shoulder the burden of adaptation. This paper investigates how visual occlusion of the object being passed affects the subjective perception and quantitative performance of the human receiver. We performed an experiment in virtual reality where seventeen participants were tasked with repeatedly reaching to take a tool from the hand of a robot; each of the three tested objects (hammer, screwdriver, scissors) was presented in a wide variety of poses. We carefully analysed the user's hand and head motions, the time to grasp the object, and the chosen grasp location, as well as participants' ratings of the grasp they just performed. Results show that initial visibility of the handle significantly increases the reported holdability and immediate usability of a tool. Furthermore, a robot that offers objects so that their handles are more occluded forces the receiver to spend more time in planning and executing the grasp and also lowers the probability that the tool will be grasped by the handle. Together these findings indicate that robots can more effectively support their human work partners by increasing the visibility of the intended grasp location of objects being passed.**

***Index Terms*—Human-Robot Interaction; Human Grasping; Object Handovers; Visual Occlusion; Virtual Reality**

I. Introduction

Humans grasp objects to perform many types of skillful manipulation, such as using a screwdriver to secure a door hinge to a cabinet. The choice of grasp position on a tool is crucial to perform the task efficiently and succeed at the overarching goal. Robots are increasingly expected to have similarly adaptive and intelligent skills in grasping and manipulation. In the context of Industry 4.0 [1], robots are envisioned to work alongside and actively interact with human workers in order to improve working conditions by decreasing physical labour fatigue and increasing safety. Fields ranging from construction and manufacturing to healthcare and eldercare all stand to benefit from efficient interactions

* These authors contributed equally to this work.

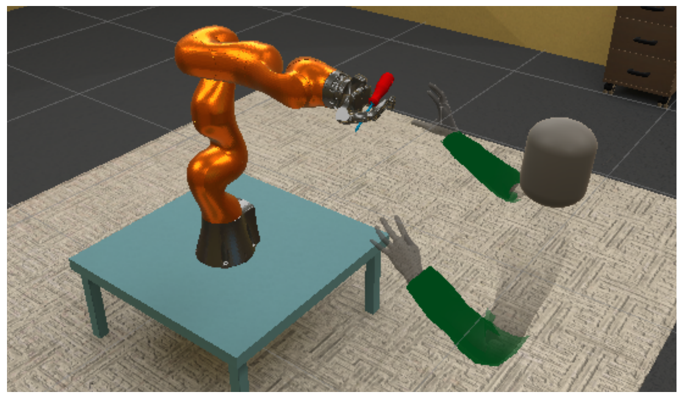

Fig. 1. Example of a scene in our VR environment. A KUKA LBR iiwa robot offers a screwdriver to the human participant. The user's hands, arms, and head are tracked with a headset and are depicted in gray and green.

between humans and robots [2]. In particular, robots need to be able to collaborate efficiently with human workers on physical tasks, *e.g.*, fetching and handing over a tool [3]. A robot-to-human handover requires an object-offering algorithm that enables the human to receive a tool from the robot and then use it. The appropriate presentation of such objects is likely to be crucially important to the success of the collaboration.

Different grasps by a robot can dramatically affect the grasp strategy of the human partner [4], impacting quantitative measures such as grasping time and subjective indices such as grasping confidence. This work investigates *how visual occlusion of the object being passed by a robot influences the quantitative grasp planning and reaching performance and the subjective evaluation by the human receiver*. To this end, we performed an experiment in virtual reality (VR) where a simulated robot offers tools with different degrees of visual occlusion (Fig. 1). The participant is instructed to reach to take the object from the robot to perform an object-specific task, such as driving a screw with a screwdriver. Once the grasp is completed, participants report how well they would be able to hold on to the object (holdability) and how well they would be able to perform the task without any further re-manipulation (immediate usability). Analyses of hand motion, head motion, reaction time, grasping time, grasp placement on the object, and subjective responses indicate how participants change their grasping strategy depending on the visual occlusion of the tool handle. This work is a step toward enabling robots to choose an appropriate grasp for a given handover by considering visual occlusion and its effects on the human receiver.

II. Related Work

A considerable amount of research in behavioural science shows that humans coordinate their motions to achieve a common goal during joint actions such as object handovers [5]–[8]. In other words, humans tend to plan their motions by considering their partner's needs, representing their partner's actions, and predicting the consequent outcomes [9], [10]. For this reason, scientists argue that humans form shared representations of the task to better predict each other's movements and to act accordingly [11]. Efficiency and social cohesion are also listed as reasons to adopt such shared representations. From this perspective, the goal of a passer's grasp on an object is not to retain the object, but rather to pass it so that the receiver can effectively perform a certain task [3], [12].

A robot should consider similar factors in order to be accepted and appreciated as a partner, and to positively influence the performance of the human it is supporting. It has recently been shown that a robot using different grasps while passing objects impacts the performance of a human receiver [4]. In particular, a robot grasp strategy that accounts for the subsequent task of the human receiver can reduce the time to complete the task by eliminating post-handover readjustments of the object [4]. Moreover, human perceptions of the interaction improve especially when the use of the object introduces strict constraints on the choice of grasp [4]. Therefore, the robot grasping strategy is pivotal for the success of a handover and should account for the receiver's task.

In general, a successful robot grasp is characterised by stability [13] and/or speed [14]; the task to perform [15] and its requirements in terms of force and mobility [16], [17] are often ignored. In detail, many aspects influence the choice of a grasp [18]–[21], including the initial pose of the object; the environmental constraints, such as gravity; the object constraints such as shape, size and function; the task constraints, such as the need to place the tip of the screwdriver on the screw and to twist the handle to apply enough torque; the gripper constraints such as the size of the robot hand relative to the object; and the habits of the agent such as experience and social convention. While there is no consensus as to which aspect is predominant for a grasp choice, and much work has investigated subsets of these aspects, there is general agreement that grasping is a purposive action [22] and is situation-dependent [17], [23]. Such considerations are important for robots to choose efficient and effective behaviours when collaborating with humans. In situations where robots hand objects (such as tools) to humans, the object should be easily accessible to the human. This goal poses additional constraints on the grasp selection for the robotic system, as the areas that are easier to grasp by humans may overlap with areas where a large number of stable grasps can be found for the robotic gripper. For instance, the handle of a knife is likely to be the preferred area for grasping by both human and robot, but selecting a handle grasp that maximizes the ease of execution for a robotic system may increase the time needed for the human to grasp the object and/or encourage potentially aversive or harmful human actions such as grasping a part of the blade.

Presented object orientation and robot pose are known to be critical to the performance of a handover [24]–[26]. However, humans can react only to their estimates of these physical quantities, which must be acquired via visual perception. When an object's handle is clearly visible, its position and orientation are easy to perceive. Thus, we decided to focus our attention to how visual occlusion impacts the planning of a grasp and the subjective perception of the handover. Our focus constitutes a fundamental difference to [4], who concentrated on post-handover manipulations and task completion time. To our knowledge, this is the first study on the influence of visual occlusion of the object on the performance of a human receiver. We do not focus on the success rate of the object exchange, as we postulate that humans can easily leverage their manipulation and motion skills to succeed in grasping well-known objects. In contrast, we focus on the planning and the perception of the receiver's reach as a function of the presentation of the object. Specifically, we focus on visual occlusion of the handle because it is an observable quantity that could directly impact the receiver.

Generally, testing physical human-robot interaction (HRI) approaches necessitates conditions that are both repeatable and similar to real-world scenarios. However, cost, time, safety, available equipment, and natural physical variability can impose limitations that prevent full repeatability. Many of these shortcomings can be overcome using virtual reality (VR), whose reliability for testing HRI approaches has been positively assessed [27]. This study leverages VR to ensure full repeatability of each object offering across participants and to enable the testing of a higher number of handover scenarios in a shorter period of time. Of note, we do not try to simulate physical interactions between the receiver's hand and the object; instead, we use VR to analyse planning and reaching to grasp. VR has been shown to be effective in human object manipulation and grasping even without haptic feedback [28], [29]. The use of VR, augmented reality (AR), and mixed reality (MR) has gained momentum as an alternative platform for conducting interaction studies in HRI, with several works replicating non-virtual studies and testing collaboration scenarios [30]–[32]. Moreover, VR, AR, and MR are increasingly used in applications with and for robotics [33].

III. Materials and Methods

A. Experimental Setup and Protocol

We developed a VR experiment using Unity (Unity Technologies, 2019.3.11f1) and an Oculus Quest VR headset

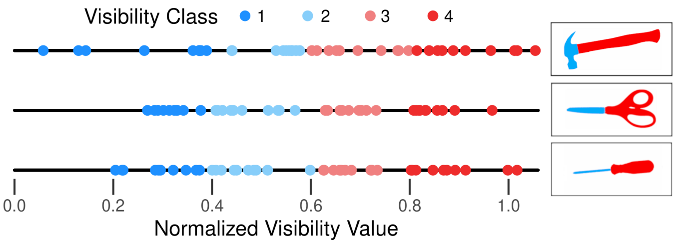

Fig. 2. Normalised visibility values and visibility class for all trials presented in the experiment. Each point represents one trial for the corresponding object. The pictures on the right show the images used to calculate the approximate maximum handle visibility for each object.

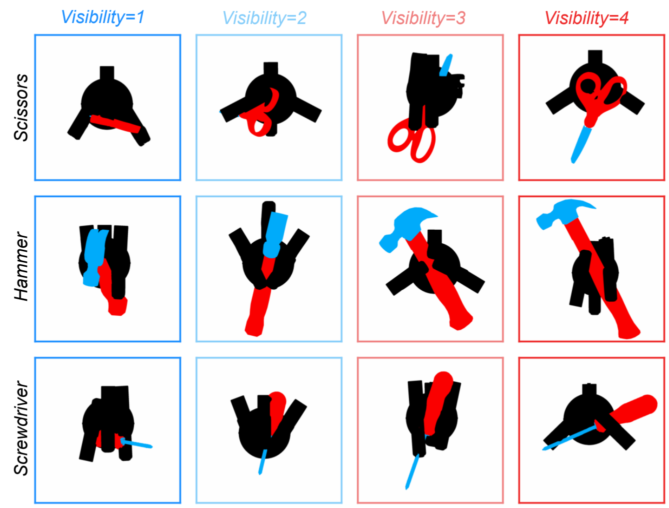

Fig. 3. Sample grasps tested in the study. Each column corresponds to a visibility level, and each row shows one object.

(Facebook Technologies, USA), which provides controller-free tracking of the user's hands and fingers at 60 Hz and visual rendering of the 3D scene at 90 Hz. The experimental scene consisted of a fixed-base KUKA LBR iiwa robot arm located on a table in the center of a room and equipped with a three-fingered gripper (BarrettHand BH8-262) (Fig. 1). The participant sat on a chair with no wheels and no armrests, and they could move their upper body freely. In each trial, the robot offered one of three tools taken from the YCB dataset [34]: hammer, screwdriver, and scissors. The pose of the robot arm was kept constant throughout all trials so that the robot hand was always in the same position. The pose of the fingers of the robot hand varied across simulated grasps. The handover location was therefore constant throughout the study.

The objects were offered at varying levels of visual occlusion of the area designed for grasping, *i.e.,* the handle (Fig. 2). We defined *visual occlusion* as the number of handle pixels visible from a default user viewpoint divided by the number of handle pixels visible from this same viewpoint when the object is held in the iconic highly visible pose shown in Fig. 2. This definition yields a continuous visibility score ranging from 0 to slightly more than 1; a small number of grasps have a normalized visibility value greater than 100% because the handle was even more visible than in the hand-selected iconic pose. This continuous score was further divided into four visibility levels. These approximate visibility categories allowed us to balance the experimental conditions and to use a block experimental design.

We used a Unity-based grasp simulator called PrendoSim [35] (https://prendosim.github.io) to generate over 1000 physically realistic grasps for each of the three objects by randomly changing the initial pose of the object with respect to the robot hand. Though it varied substantially, the orientation of the object was not explicitly controlled in this experiment. Given the low number of stable grasps found for visibility less than 20%, the four visibility levels were defined as level 1: $<$40%, level 2: 40–60%, level 3: 60–80%, and level 4: $>$80%. We randomly picked ten grasps for each level of visibility for each object, yielding 120 trials in total, *i.e.*, 3 objects $\times$ 4 levels of occlusion $\times$ 10 scenes. Fig. 2 reports the handle visibility for all selected trials, and Fig. 3 shows cartoon versions of twelve sample grasps used in the study. The trial order was randomised for each participant.

At the beginning of the experiment, we collected demographics and handedness information. The participant then took part in a training session by performing grasps to familiarise themselves with the experimental protocol. Users were instructed to grasp each tool as if planning to use it. In particular, they were asked to grasp the hammer to hammer in a nail, the screwdriver to drive a screw, and the scissors to cut a piece of paper. A VR animation showed how each tool would be used. The sequence of a trial is described visually in Fig. 4, and this paper's supplementary video shows several sample trials. At the start of each trial, the participant placed their hands inside highlighted spheres near their waist to ensure consistency in initial conditions across trials. When the robot appeared, the starting spheres disappeared, and the participant could start reaching to grasp the object with their right hand. As soon as the hand of the participant virtually touched the object, the reach was considered finished and the scene dissolved. If a participant contacted the robot while performing this task, the collision was ignored. After each trial, participants answered two questions. Question 1 (Q1) was: *Given the grasp you just performed, how well would you be able to hold the object?* Question 2 (Q2) was: *Given the grasp you just performed, how easy would it be to use the object immediately?* The participant answered these questions on a visual analogue scale, where the minimum meant "Not at all" and the maximum meant "Extremely".

We wanted participants to think about the task to perform with the tools; thus, we introduced another type of trial in which they were asked to grasp the object as fast as possible. These trials were signalled with a "Timed" sign. Each participant experienced 30 timed trials interspersed randomly with the 120 task trials. Therefore, the experiment consisted of 150 trials, which were divided into 10 sets of 15 trials. Participants could take a break at the end of each set and independently decide the length of such breaks. They could also discard a trial whenever they felt that something unexpected had

Fig. 4. Sample trial in our experiment. Top row, left to right: The participant waits with their hands in the designated spheres. Then they reach for the object presented by the robot and grasp it; the detected contact points are highlighted in cyan but were not visible during the study. Bottom row, left to right: The user answers Q1 and Q2 about the grasp they just performed. The final picture shows an aggregate of the contact points of multiple users for this scene.

occurred that might have affected their performance, such as a tracking error or a lapse in attention. These data were then marked as invalid and not used in the analyses. Overall, the experiment took 50 to 90 minutes to complete.

Conducting this study in VR allowed us to achieve accurate repeatability of the trials, presentation of a high number of scenes, accurate hand and head tracking, and automatic placement of the object in the robot hand at the start of each trial. These four benefits would have been difficult to achieve with a real robot. However, using VR also has drawbacks: it may not be analogous to the real-world setting, and participants may respond differently. We tried to address these issues by including realistic details to enable a high degree of immersiveness, and we terminated each trial when the user touched the object to avoid a mismatch between visual and haptic feedback. Stopping at contact thus preserved the illusion of kinematic interactions (reaching), and allowed us to focus on grasp planning and reaching. Furthermore, we allowed the participants to take a break and root back to the physicality of reality after completing each set of trials.

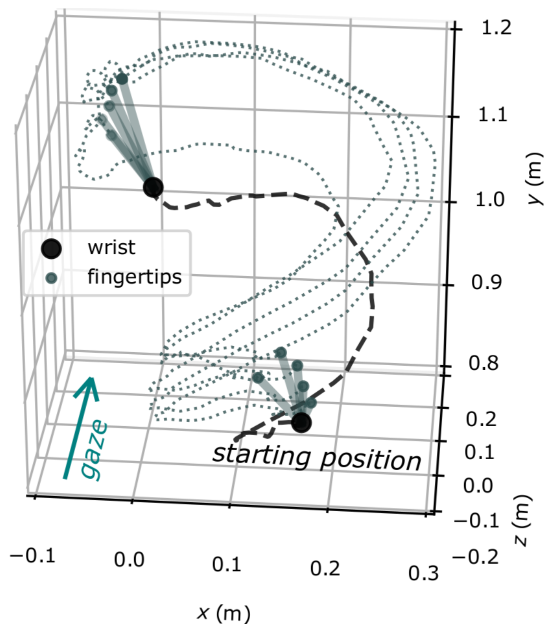

Fig. 5. Trajectory of the user's right hand from a sample experimental trial.

B. Data Analysis

In each trial we collected the trial number, trial scene, reaction time, grasp time, wrist position over time, head position over time, and answers to Q1 and Q2. Fig. 5 shows a sample trial from the study. The reaction time is defined as the time between the start of the trial and when the hand of the participant left the highlighted starting sphere. The grasping time is defined as the time between that instant and the first touch with the object, *i.e.*, the total trial time minus the reaction time. The motion signals (the position of the right wrist over time and the position of the head over time) were first smoothed with a second-order Butterworth low-pass filter with a cutoff frequency of 2 Hz to remove motion artifacts and outliers caused by Quest tracking errors. Then the length of each right-hand trajectory was computed as the sum of finite differences between two consecutive time instants over the entire trial time. We defined the hand excursion as the difference between the actual hand trajectory length and the minimum distance between the starting and ending hand positions in a straight line. The length of each recorded head trajectory was computed in the same way as the hand trajectory. Finally, to distinguish lateral motions that provide a new viewing angle from normal motions that increase the viewed size of the object, we defined the lateral head motion as the length of the head trajectory projected onto the frontal plane of the user.

We analysed changes in behavioural variables with increasing object visibility levels by assessing the impact of visibility on all six of the calculated metrics. To this end, we used generalized linear mixed-effects models (GLMM, via the lme4 R package) to predict the corresponding target variable and model intercept for each participant. GLMMs are a generalization of linear mixed-effects models that extends to dependent variables sampled from non-normal distributions such as reaction time data [36]. To assess the impact of visibility on grasp location (*i.e.*, whether the object was grasped by the handle) we used a one-tailed two-proportion Z-test to detect whether the proportion of grasps occurring on the handle increased with visibility level. Finally, we used linear mixed-effects models to measure the impact of visibility on participants' responses to Q1 and Q2. Throughout our statistical analyses, we used a likelihood ratio test to determine whether visibility has a significant effect on each measured outcome, and we used $\alpha = 0.05$ as the level at which we reject the null hypothesis.

C. Participants

Seventeen participants took part in the experiment (gender: female = 10, male = 7; age $\in [26, 58]$ years, μ_{age} = 32.8, σ_{age} = 7.9). All participants were healthy, were right-handed (16) or ambidextrous (1), reported normal or corrected-to-normal vision, reported no history of neurological diseases, and had unencumbered movement of their right hand. Participants were not aware of the purpose of the experiment. The study was advertised via mailing lists and Facebook posts. Individuals participated on a voluntary basis and gave their signed consent at the start of the study. The experiment followed procedures approved by Haptic Intelligence Department's framework agreement with the Max Planck Ethics Council, with protocol number F016A.

IV. Results

A. Task versus Timed Trials

We compared how participants performed task and timed trials by examining reaction time and grasp time, regardless of visibility conditions, as reported in Fig. 6. If the trial type affected participants' actions, we would expect to see higher time metrics in task trials compared to timed trials. We used a Wilcoxon signed rank test to compare the values of each metric averaged for each participant. Our analyses revealed that reaction times and grasp times were indeed higher for the task trials ($V = 138, p < 0.01$ and $V = 146, p < 0.001$, respectively, with $n = 17$). These results indicate that user behaviour in task trials differed from that in timed trials, as intended.

B. Visibility Effects on Behavioural Outcomes

We assessed the impact of visibility on our behavioural metrics in task trials. Fig. 7 and Table I show these results.

1) Reaction Time: Reaction time is modelled as the dependent variable, visibility level as the fixed effect factor, and participant ID and experimental block as random effects. Reaction time data are specified to be sampled from a gamma distribution with an identity link function. A significant effect of visibility on reaction time was observed ($\beta = -0.04$, $p < 0.0001$), indicating a reduction of reaction time with increasing visibility; in other words, grasping low-visibility tools requires more planning time.

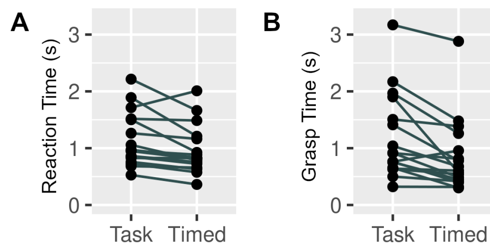

Fig. 6. Average A) reaction time and B) grasp time for each participant in the task and timed trials.

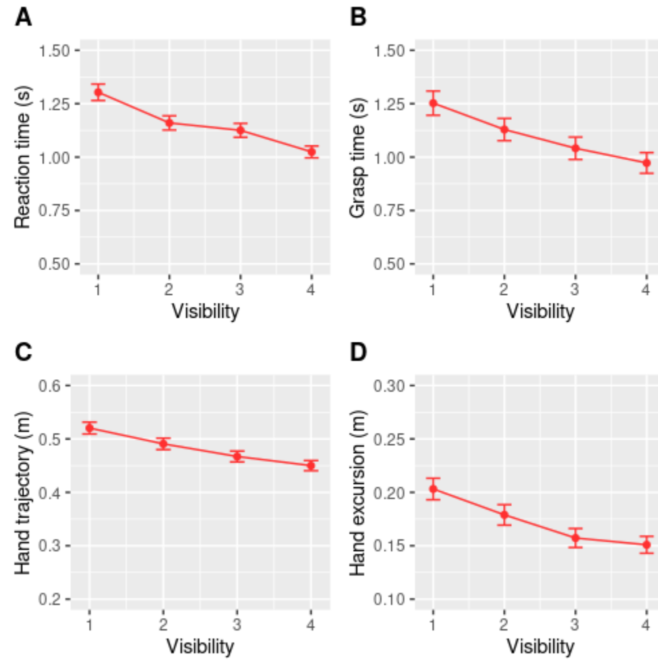

Fig. 7. Mean and standard error of A) reaction times, B) grasp times, C) hand trajectories, and D) hand excursions across all users for visibility levels 1–4.

2) Grasp Time: We use the same framework to analyse the effect of visibility on grasp time. A significant effect of visibility on grasp time was observed ($\beta = -0.06$, $p < 0.0001$), indicating that participants needed less time to move to grasp the tool when the handle visibility was high.

Variable	β	95% C.I.	Significance
Reaction time (s)	–0.04	–0.06, –0.03	$\leq$ 0.0001
Grasp time (s)	–0.06	–0.09, –0.03	$\leq$ 0.001
Hand trajectory (m)	–0.02	–0.03, –0.01	$\leq$ 0.0001
Hand excursion (m)	–0.01	–0.011, 0.005	$\leq$ 0.0001
Head motion (m)	–0.01	–0.013, –0.004	$\leq$ 0.0001
Lateral head motion (m)	–0.018	–0.023, –0.013	$\leq$ 0.0001

TABLE I
Fixed-effect coefficients for the effect of visibility on behavioural outcomes.

3) Hand Motion: Next we inspect the amount of hand movement recorded under different visibility conditions. There was a significant effect of visibility on the length of the hand trajectory, with a reduction observed with increasing visibility levels ($\beta = -0.02$, $p < 0.0001$). However, the confidence intervals are wide relative to the size of the effect, indicating that the effect may vary. To better assess the effects of handle visibility on hand movement, we examined hand excursion, which is the extra distance the hand travelled to reach the tool, compared to the shortest path. There was a significant effect of visibility on hand excursion; the participant's hand trajectory was typically closer to the minimum when the handle was more visible ($\beta = -0.1$, $p = 0.0001$).

4) Head Motion: Next we inspected the amount of head movement participants made under different visibility conditions. There was a small but significant decrease in the total length of the head trajectory associated with object handle visibility ($\beta = -0.01$, $p < 0.0001$). We performed the same analysis on lateral head motion; results show that lateral head motion decreased as visibility improved ($\beta = -0.018, p \leq 0.0001$). Taken together, these findings indicate that lower levels of handle visibility caused users to move their head laterally so they could observe the object from multiple angles.

C. Grasp Location

For task trials, the participants were instructed to grasp the object as if they intended to use it in the relevant task (hammering, screwing, or cutting). In optimal circumstances, where the object handle was clearly visible, we would expect the participant to grasp the tool by the handle. To determine whether visibility affected grasp location, we performed a two-proportion Z-test to detect whether the proportion of grasps occurring on the handle increased with visibility level; we divided each object into the red handle zone and blue non-handle zone shown in Fig. 2, and we calculated the average proportion for each participant in each visibility level. The Z-test estimates the likelihood of two measured proportions arising due to chance. We used a directional hypothesis, as we expect more grasps on the object handle for higher visibility levels, and compared adjacent visibility levels in pairs. The data are shown in Fig. 8. There was a statistically significant increase in grasps occurring on the handle when comparing visibility levels 1 and 2, with $Z = 4.12, p \leq 0.0001$, and levels 3 and 4 with $Z = 4.33$, $p \leq 0.0001$. The increase between levels 2 and 3 is not statistically significant ($Z = 1.17$, $p = 0.12$).

D. Reported Holdability and Usability

Finally, we analysed the effect of visibility on responses to the two questions that followed each trial, which assessed the subjects' self-reported ability to *hold* and *use* the object they had just reached to grasp. These results are shown in Table II and in Fig. 9, using 0 for the minimum response and 100 for the maximum response.

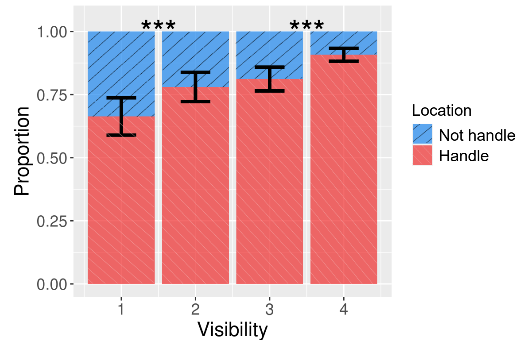

Fig. 8. Proportion of grasps occurring on the handle of the object by visibility level. The average proportion is computed for each participant. Error bars indicate the standard error across participants, and asterisks indicate statistically significant increases of grasps occurring on the handle when comparing adjacent visibility levels.

	Variable	β	95% C.I.	Significance
Q1	Visibility	9.12	7.68, 10.56	≤ 0.0001
	Object : screwdriver	–14.85	–20.42, –9.27	≤ 0.0001
	Visibility : scissors	–3.76	–5.78, 1.73	≤ 0.0001
Q2	Visibility	13.23	11.44, 15.02	≤ 0.0001
	Object : screwdriver	–7.92	–14.83, –1.01	≤ 0.05
	Object : scissors	20.77	13.86, 27.68	≤ 0.0001
	Visibility : scissors	–8.74	–6.23, 0.07	≤ 0.0001

TABLE II
FIXED-EFFECT COEFFICIENTS FOR THE LINEAR EFFECT OF VISIBILITY AND OBJECT INTERACTION. ONLY STATISTICALLY SIGNIFICANT COMPARISONS ARE REPORTED.

1) Q1: "Given the grasp you just performed, how well would you be able to hold the object?": We analysed the responses to the questions using linear mixed effects models, with the recorded responses as the outcome variable. We constructed several models of increasing complexity and performed a likelihood ratio test to test for effects contributed by each predictor. We thus started with including only the random effect of participant as a predictor, and we consecutively included a fixed effect of visibility and a random effect for block repetition to account for both effects of interdependence of measurements within each subject, as well as fatigue and practice. Finally we included an interaction term for objects. There was a significant effect of visibility on the Q1 response, with $\beta = 9.1163$ and $p \leq 0.0001$. The responses for trials with the screwdriver are significantly lower compared to the hammer ($\beta = -14.85$, $p \leq 0.001$). The effect of visibility on the holdability of the scissors was smaller compared to the hammer ($\beta - 3.76$, $p \leq 0.001$).

2) Q2: "Given the grasp you just performed, how easy would it be to use the object immediately?": Repeating the same procedure for the second question, we observed a significant fixed effect of visibility on the response ($\beta = 13.23$, $p \leq 0.0001$). Usability responses on screwdriver trials were found to be lower than on hammer trials ($\beta = -7.92$, $p \leq 0.05$). The responses for trials with scissors were found to be significantly higher than for the hammer ($\beta = 20.77$, $p \leq 0.0001$); however, as visibility increased, its effect on the Q2 response for scissors diminished ($\beta = -8.74, p \leq 0.0001$).

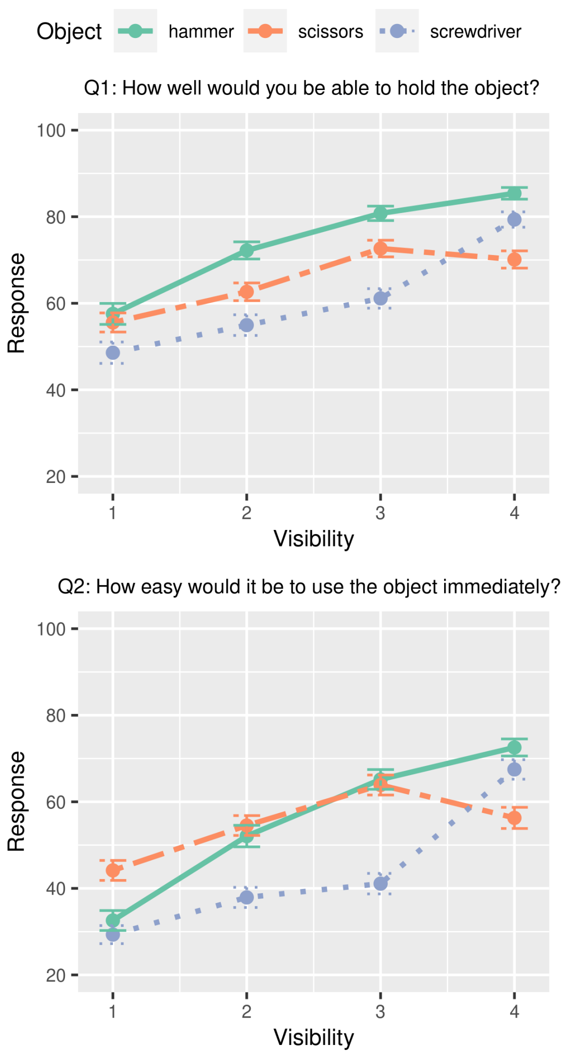

Fig. 9. Mean response values and standard errors across users for Q1 and Q2, separated by object and visibility level.

V. Discussion

A. *Handle Visibility Positively Affects Behaviour, Holdability, and Usability*

Our results indicate that participants were generally faster in planning and reaching to grasp the presented tool when its handle was more visible. Moreover, participants were more likely to grasp the handle, and they had higher confidence at being able to hold and successively use the object given the grasp they performed. These results identify handle visibility as a positive factor to the success of an object handover and visual occlusion as a negative factor affecting the behaviour and subjective perception of the receiver.

Our experiment did not test the effect of visual occlusion on the success of using the object since a trial was considered finished once the grasp had occurred. Thus, we further interpret our results of lower subjective ratings being associated with lower visibility as an expression of user frustration toward the lack of cooperation by the robot. Anecdotally, some participants spoke to the robot in disbelief on trials with low visibility. These findings align with [4] in that humans adapt well to sub-optimal object offers with a severely occluded handle at the cost of performing compensatory grasping strategies and post-grasp re-adjustments.

We noticed two contrasting behaviours in the task trials, where participants were supposed to grasp the tool in order to use it immediately. Participants tended either (1) to aim at grasping the handle or (2) to grasp the part that was closest or most available, which compromises their ability to directly perform the task. In case (1), participants seemed to feel a strong bias due to the instruction of grasping to use the object, and they preferred grasping the object on its intended grasping affordance. This choice potentially necessitated extra time and entailed a general degradation of holdability and possibly of usability when the object was severely occluded. In case (2), participants exhibited a more flexible behaviour. They seemed to prefer to grasp the handle; however, when its occlusion was severe, they opted for grasping the tool's non-handle part, thus prioritising holdability over direct usability. We believe that this second behaviour is motivated by the excellent human skills in manipulation that easily permit in-hand or post-grasp re-manipulations. These results on the effects of visual occlusion are in line with previous results in [4], where it was noted that participants preferred a temporary grasp when the grasping affordance was physically occluded. Our study extends this prior knowledge by providing strong evidence that the *visibility* of an offered tool's handle greatly facilitates task-oriented handovers from robot to human.

B. *Definition of Visual Occlusion and its Ramifications*

The scenes in this study were produced using the Unity-based simulator PrendoSim [35]. The object's level of occlusion was computed by assuming a fixed point of view and hand-picking a highly visible reference pose for each object. The rationale behind this choice is that a robot often does not know or cannot foresee the exact physical characteristics of its human partner, such as eye height, gaze direction, or visual acuity. We thus decided on a fixed point of view in front of the robot at medium height. Human beings are generally able to move around quickly; our participants had different heights and moved their trunk and head during the experiment, thus acquiring additional visual information from novel points of view (as measured by our head motion metrics). Our results indicate that even with some change in perspective through head motion, low levels of handle visibility computed from our chosen fixed point of view cause lower user confidence in holding and subsequently using the object. Another reason behind our choice of fixed point of view derives from situations where the human partner cannot move freely and/or cannot be easily observed by an assistive robot, as is the case for a mechanic working under a car [37]. In such scenarios, it would be appropriate to estimate an approximately constant position of the mechanic's head when offering tools.

Visual occlusion is generally caused by physical occlusion and the pose of the object with respect to the robot hand

and the user's head. While interdependent, these quantities are fundamentally different. Physical occlusion implies a certain level of visual occlusion, as the hand of the robot is covering a part of the surface of the object, thus increasing its visual occlusion. However, visual occlusion does not necessarily entail physical occlusion. For example, when the hammer is grasped on its head and the handle is oriented back toward the body of the robot, the handle is visually occluded but not physically occluded. This relationship can be expressed logically as $occ_{phys} \Rightarrow occ_{vis}$ and $occ_{vis} \not\Rightarrow occ_{phys}$. Our consideration of visibility rather than physical exposure adds a new layer to the capabilities robots could soon demonstrate.

The orientation of the object with respect to the user also certainly plays an important role in how the offered object is perceived [25], [26]. Orientations of the handle angled toward the partner are usually preferred, as they offer more unencumbered area and seem socially more natural, as if the robot is really offering the object to the receiver. Laterally oriented offerings suffer from a strong dependence on the handedness of the receiver; an object handle aimed to the side of the grasping hand is typically easier to reach than a handle in the opposite direction. Finally, if the object handle is offered oriented away from the human, *i.e.*, toward the robot, then the object is more visually occluded and the offering is perceived as awkward, as if the robot is not ready or willing to collaborate. Post-hoc visual inspection of the object poses presented in our study shows that the raw rotation values and visibility values vary widely relative to one another but are not completely statistically independent; they are related by the physics of grasping and the geometry of the viewing scene. However, when we added object orientation values to our statistical models of participant responses, we saw that the effect of orientation was either insignificant or very small compared to visibility. Thus, we believe the influence of visual occlusion on handovers is somewhat distinct from that of object orientation. Indeed, the visibility of the handle decreases as it is rotated toward the user because it becomes aligned with the viewing axis. The pose in which an oblong object has maximum visibility is angled away from the receiver. Optimal tool offering may require balancing of these quantities; future research should systematically investigate their interplay.

C. Implications for Human-Robot Interaction

We believe that introducing consideration of visual occlusion in the grasp selection process will enable a robot to choose a grasp that will improve its interaction with the human receiver, and that will require the human receiver to adapt less to the robot. In particular, a useful ad hoc rule could be to offer the grasping part of the object so that it is more than 80% visible with respect to its maximum visibility, as being above this value (visibility level 4) significantly increases the chances of the receiver grasping the tool's handle (Fig. 8). Following this rule would positively impact the perceived holdability and direct usability of the object after the handover. Future work should investigate and refine this rule of thumb to confirm its positive effects on human users.

Whenever it is feasible for a robotic agent to approximate the point of view of its human counterpart a priori, a robot could also use metrics such as reaction time as feedback to assess the quality of its offering of objects. This strategy would allow the agent to learn and adapt to the human user's preferences without the need for explicit feedback by the user.

D. Limitations and Directions for Future Work

This work focused on visual occlusion; however, physical occlusion and the orientation of the object are also crucial in the planning of a grasp. Future work should aim to better understand the relationship among these three intertwined factors. The position of the object and the pose of the robot arm should also be varied in future work to better approximate realistic handover scenarios. Finally, we believe that continuing to display the performed grasp after it is complete could increase the quality and reliability of the user responses in such studies.

Another important observation can be made about the impact of the object on the effect of visual occlusion. Our analyses revealed that the effect of visibility on participant responses was not uniform across all objects. Specifically, the ratings for the screwdriver were lower on both questions, the hammer was rated highest overall, and the benefits of visibility for scissors plateaued at higher visibility levels. The effects of visual occlusion may be connected to the severity of the kinematic constraints that must be followed to use each object, as well as their overall size. One possible interpretation of our results is that higher levels of object-use constraints reduce the amount by which visibility affects the user's ability to perform a task-oriented grasp. As an example, when grasping the highly constrained scissors, one is likely to first grasp the tool with a temporary grasp that would not allow direct use, and then to rearrange one's grasp for the task. This case exemplifies how the visual occlusion of the object might not be the determining factor when objects have strong use constraints. From this perspective, future work should investigate a wider pool of objects and tools in order to further understand the correlation between use constraints and how visual occlusion affects grasp performance.

Finally, it would be valuable to investigate human grasping behaviour when facing unknown objects such as geometrically complicated antagonistic objects whose use is not intuitive, or objects designed to be used in multiple manners, thus investigating how visual occlusion impacts such choice. Another version of this experiment could have the robot pass the wrong tool, thus forcing the receiver to think about situational affordances, as in [38].

Acknowledgments

The authors thank Eric Young, Natalia Egana Rosa and Maria-Paola Forte for their critical feedback as pilot participants. This research was partially supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2120/1 – 390831618.

References

[1] E. H. Østergaard, "The role of cobots in industry 4.0," https://info.universal-robots.com/hubfs/Enablers/Whitepapers/Theroleofcobotsinindustry.pdf, 2017, (Accessed: May 25, 2020).

[2] A. Ajoudani, A. M. Zanchettin, S. Ivaldi, A. Albu-Schäffer, K. Kosuge, and O. Khatib, "Progress and prospects of the human-robot collaboration," *Autonomous Robots*, 2017.

[3] V. Ortenzi, A. Cosgun, T. Pardi, W. P. Chan, E. Croft, and D. Kulić, "Object handovers: A review for robotics," *IEEE Transactions on Robotics*, pp. 1–19, 2021.

[4] V. Ortenzi, F. Cini, T. Pardi, N. Marturi, R. Stolkin, P. Corke, and M. Controzzi, "The grasp strategy of a robot passer influences performance and quality of the robot-human object handover," *Frontiers in Robotics and AI*, vol. 7, p. 138, 2020.

[5] N. Sebanz, H. Bekkering, and G. Knoblich, "Joint action: Bodies and minds moving together," *Trends in Cognitive Sciences*, 2006.

[6] D. A. Rosenbaum, K. M. Chapman, M. Weigelt, D. J. Weiss, and R. van der Wel, "Cognition, action, and object manipulation," *Psychological Bulletin*, vol. 138, no. 5, pp. 924 – 946, 2012.

[7] S. Endo, G. Pegman, M. Burgin, T. Toumi, and A. M. Wing, "Haptics in between-person object transfer," in *Haptics: Perception, Devices, Mobility, and Communication. EuroHaptics 2012. Lecture Notes in Computer Science*, P. Isokoski and J. Springare, Eds. Springer, Berlin, Heidelberg, 2012, vol. 7282, pp. 103–111.

[8] M. Controzzi, H. Singh, F. Cini, T. Cecchini, A. Wing, and C. Cipriani, "Humans adjust their grip force when passing an object according to the observed speed of the partner's reaching out movement," *Experimental Brain Research*, vol. 236, no. 12, pp. 3363–3377, Dec. 2018.

[9] D. M. Wolpert and Z. Ghahramani, "Computational principles of movement neuroscience," *Nature Neuroscience*, 2000.

[10] L. M. Sacheli, E. Arcangeli, and E. Paulesu, "Evidence for a dyadic motor plan in joint action," *Scientific Reports*, 2018.

[11] M. Ray and T. N. Welsh, "Response selection during a joint action task," *Journal of Motor Behavior*, vol. 43, no. 4, pp. 329–332, 2011.

[12] F. Cini, V. Ortenzi, P. Corke, and M. Controzzi, "On the choice of grasp type and location when handing over an object," *Science Robotics*, vol. 4, no. 27, 2019.

[13] A. Bicchi and V. Kumar, "Robotic grasping and contact: a review," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2000.

[14] J. Mahler, R. Platt, A. Rodriguez, M. Ciocarlie, A. Dollar, R. Detry, M. A. Roa, H. Yanco, A. Norton, J. Falco *et al.*, "Guest editorial open discussion of robot grasping benchmarks, protocols, and metrics," *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 4, pp. 1440–1442, 2018.

[15] J. Bohg, A. Morales, T. Asfour, and D. Kragic, "Data-driven grasp synthesis – a survey," *IEEE Transactions on Robotics*, vol. 30, no. 2, pp. 289–309, 2014.

[16] T. Pardi, R. Stolkin, and A. M. Ghalamzan, "Choosing grasps to enable collision-free post-grasp manipulations," in *Proceedings of the IEEE-RAS International Conference on Humanoid Robots*, 2019.

[17] 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.

[18] J. Landsmeer, "Power grip and precision handling," *Annals of the Rheumatic Diseases*, vol. 2, pp. 164–70, 06 1962.

[19] N. Kamakura, M. Matsuo, H. Ishii, F. Mitsuboshi, and Y. Miura, "Patterns of static prehension in normal hands," *American Journal of Occupational Therapy*, vol. 34, no. 7, pp. 437–445, 1980.

[20] M. R. Cutkosky, "On grasp choice, grasp models, and the design of hands for manufacturing tasks," *IEEE Transactions on Robotics and Automation*, vol. 5, no. 3, pp. 269–279, 1989.

[21] T. Iberall, "Human prehension and dexterous robot hands," *The International Journal of Robotics Research*, vol. 16, no. 3, pp. 285–299, 1997.

[22] J. R. Napier, "The prehensile movements of the human hand," *The Journal of Bone and Joint Surgery. British volume*, vol. 38, no. 4, pp. 902–913, 1956.

[23] C. Ansuini, M. Santello, S. Massaccesi, and U. Castiello, "Effects of end-goal on hand shaping," *Journal of Neurophysiology*, vol. 95, pp. 2456–2465, 2006.

[24] M. Cakmak, S. S. Srinivasa, M. K. Lee, S. Kiesler, and J. Forlizzi, "Using spatial and temporal contrast for fluent robot-human hand-overs," in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2011.

[25] J. Aleotti, V. Micelli, and S. Caselli, "An affordance sensitive system for robot to human object handover," *International Journal of Social Robotics*, vol. 6, no. 4, pp. 653–666, 2014.

[26] W. P. Chan, M. K. X. J. Pan, E. A. Croft, and M. Inaba, "Characterization of handover orientations used by humans for efficient robot to human handovers," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015.

[27] V. Villani, B. Capelli, and L. Sabattini, "Use of virtual reality for the evaluation of human-robot interaction systems in complex scenarios," in *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2018, pp. 422–427.

[28] S. Oprea, P. Martinez-Gonzalez, A. Garcia-Garcia, J. A. Castro-Vargas, S. Orts-Escolano, and J. Garcia-Rodriguez, "A visually realistic grasping system for object manipulation and interaction in virtual reality environments," *Computers & Graphics*, vol. 83, pp. 77–86, 2019.

[29] M. Furmanek, L. Schettino, M. Yarossi, S. Kirkman, S. Adamovich, and E. Tunik, "Coordination of reach-to-grasp in physical and haptic-free virtual environments," *Journal of NeuroEngineering and Rehabilitation*, vol. 16, pp. 1–14, 06 2019.

[30] L. Wijnen, P. Bremner, S. Lemaignan, and M. Giuliani, "Performing human-robot interaction user studies in virtual reality," in *Proceedings of the IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 2020, pp. 794–794.

[31] V. Weistroffer, A. Paljic, L. Callebert, and P. Fuchs, "A methodology to assess the acceptability of human-robot collaboration using virtual reality," in *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST)*. New York, NY, USA: Association for Computing Machinery, 2013, p. 39–48. [Online]. Available: https://doi.org/10.1145/2503713.2503726

[32] L. Wijnen, S. Lemaignan, and P. Bremner, "Towards using virtual reality for replicating HRI studies," in *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. New York, NY, USA: Association for Computing Machinery, 2020, p. 514–516. [Online]. Available: https://doi.org/10.1145/3371382.3378374

[33] M. Dianatfar, J. Latokartano, and M. Lanz, "Review on existing VR/AR solutions in human–robot collaboration," *Procedia CIRP*, vol. 97, pp. 407–411, 2021, 8th CIRP Conference of Assembly Technology and Systems.

[34] B. Calli, A. Singh, J. Bruce, A. Walsman, K. Konolige, S. Srinivasa, P. Abbeel, and A. M. Dollar, "Yale-CMU-Berkeley dataset for robotic manipulation research," *The International Journal of Robotics Research*, vol. 36, no. 3, pp. 261–268, 2017.

[35] D. Abdlkarim, V. Ortenzi, T. Pardi, M. Filipovica, A. Wing, K. J. Kuchenbecker, and M. Di Luca., "PrendoSim: proxy-hand-based robot grasp generator," in *Proceedings of the International Conference on Informatics in Control, Automation and Robotics (ICINCO)*. SciTePress, 2021, pp. 60–68.

[36] S. Lo and S. Andrews, "To transform or not to transform: using generalized linear mixed models to analyse reaction time data," *Frontiers in Psychology*, vol. 6, p. 1171, 2015.

[37] 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 *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2014.

[38] K. Roche and H. Chainay, "Is there a competition between functional and situational affordances during action initiation with everyday tools?" *Frontiers in Psychology*, vol. 8, p. 1073, 2017.