MASTER

The effect of joint attention on the fluency of human-robot interaction in a collaborative case picking scenario

Boonstra, Nynke

Award date:
2023

Eindhoven, 07-02-2023

The effect of joint attention on the fluency of human-robot interaction in a collaborative case picking scenario

by N. Boonstra

identity number 1251155

in partial fulfilment of the requirements for the degree of

Master of Science
in Human-Technology Interaction

Supervisors:
dr. ir. R. H. Cuijpers
MSc M. M. E. Neggers

Abstract

Combining the strengths of robots and humans often results in human-robot collaboration settings. The success of a joint action is highly dependent on the planning and anticipation skills of the agents involved. Communicating an agent's goals aids in the ability of another agent to anticipate the behavior of the first agent. One way in which humans communicate goals is by coordinating the attention of both agents to the same object of interest (joint attention). Lab research involving static humanoid robots showed that establishing joint attention (through gaze cues) can lead to a more fluent interaction and a more positive perception of the robot. The aim of the current study was to research whether this positive effect also applies to mobile non-humanoid robots in an industrial setting like collaborative case picking.

An experiment was conducted at Vanderlande to study how a stacker robot that implements different steps involved in the process of establishing joint attention, affects the (perceived) fluency of the human-robot collaboration and the perceived usefulness and ease of use of the robot. Participants were instructed to collect cases (indicated by a code presented by the robot) for the robot that stopped in front of the racks. In the first step of establishing joint attention, the robot indicated with which human order picker it intended to interact and in the second step, the robot indicated the case it intended to collect by directing attention to it. Instead of gaze cues, multi color LED lights placed on the robot and close to the cases were used to indicate what object the robot was attending to. Three different hypotheses were tested and supported by the data. The main conclusion was that a robot that indicated the target order picker was most effective in decreasing the movement time of participants and improving participants' perceptions of the robot and the collaboration. In addition, a robot that directed attention to only the target case improved only the observed fluency of the interaction. Finally, a robot that implemented both steps of a joint attention process (compared to a robot that implemented only the second step), resulted in participants perceiving the robot as significantly more useful. To conclude, the positive effect of joint attention on the (perceived) fluency of the HRC and the perceptions of the robot (in terms of usefulness and ease of use) found in static humanoid robots also applies to mobile non-humanoid robots in a collaborative case picking scenario.

Keywords: joint attention, human-robot interaction, human-robot collaboration, non-humanoid robots, mobile robots, anticipation, collaborative case picking, warehousing environment, fluency of interaction

Contents

Introduction

The interest in the development of robots that can take over repetitive and physically laborious jobs is increasing (Sharma, 2019). Robots that can completely take over the tasks originally performed by humans seem most effective, but this is not always possible. For example, robots are generally able to pick up most of the products in a warehouse, but certain products (like toilet paper) are made of a material that is hard to handle for most robots. Therefore, humans and robots need to collaborate. The aim of the current research is to study how the fluency of joint actions involving human and robotic agents can be improved. The following sections provide a literature review on joint actions in human-human interaction (HHI) and human-robot interaction (HRI). At the end, the research question and hypotheses of the current study are introduced.

Joint Actions and Anticipation

A joint action is "any form of social interaction whereby two agents or more coordinate their actions in order to pursue a joint goal" (Belhassein et al., 2022, p. 2). Planning is a key condition for successful joint action as it aids in determining how and when the agents should act (Belhassein et al., 2022; Curioni, Knoblich, & Sebanz, 2017). When agents are planning their next move, they need to anticipate (and not just respond to) the behavior of their partner so that they can adjust their own behavior effectively. This makes the interaction more fluent (Hoffman & Breazeal, 2007). For example, Mainprice and Berenson (2013) proposed a framework for planning the motions of a robot in a collaborative manipulation task with a human. They showed that taking into account the predicted human workspace occupancy (based on gesture recognition) led to safer and more efficient interactions than when only the human's current configuration was considered.

Another example of a joint action in which planning and anticipation are important is when a car driver approaches a pedestrian that intends to cross the street. The joint goal is to continue their path (preferably without stopping), while avoiding a collision. When the pedestrian notices the car, he or she might be uncertain whether the car will yield and the driver might start to slow down so he or she can stop the car in time in case the pedestrian decides to cross the street. As a result, neither the pedestrian nor the driver are sure about how to proceed, because it depends on the behavior of the other agent. A likely outcome is that both agents stop moving and wait for each other to do something. A better alternative would be for the driver to communicate in time that he or she will yield (by looking at the pedestrian and making a hand gesture), so that the pedestrian can start crossing the street earlier, which allows the car to quickly speed up again. This example shows that anticipation makes it easier to plan actions, which in turn improves the fluency of the interaction.

To anticipate the behavior of a moving object (for example, a human, an animal or a car), people often observe the current behavior. The theory of goal-directed imitation (GOADI) holds that in action imitation, and in action observation in general, the goals of an action are more important than the means with which the particular goals are being accomplished (Bekkering, Wohlschläger, & Gattis, 2000; Cuijpers, Schie, Koppen, Erlhagen, & Bekkering, 2006; Wohlschläger, Gattis, & Bekkering, 2003). This is especially the case when humans try to interpret the behavior of highly dissimilar bodies like robots or animals (Bao & Cuijpers, 2017). For example, when an octopus grasps an object with one of its arms, observers are still able to infer the target of the movement (Yekutieli, Sagiv-Zohar, Aharonov, et al., 2005; Yekutieli, Sagiv-Zohar, Hochner, & Flash, 2005). As another example, people often assume that when they see an ant dragging a dead insect across a sidewalk, it is bringing the insect back to its nest to feed other ants (Luo & Baillargeon, 2005). Attributing goals comes naturally to humans: Luo and Baillargeon (2005) found that 5-month-old infants attributed goals to a moving box (that was secretly controlled by the experimenter). To conclude, understanding observed behavior is easier when it is goal-directed. Therefore, an agent that communicates its goals (and not just its next moves), enables its human partner to better interpret and anticipate future behavior and adapt his or her behavior appropriately.

Joint Attention in Human-Human Interaction

One way in which humans communicate their goals is by establishing joint attention (Belhassein et al., 2022; Curioni et al., 2017). This means that both agents involved in the joint action coordinate their attention to the same object of interest (Bakeman & Adamson, 1984). Humans usually do this by directing their gaze towards the object or pointing at it.

To illustrate how joint attention enables people to infer goals from their partner's behavior and how this improves the fluency of a joint action, an example from daily life will be described. Imagine two friends attempting to bake pancakes together in a small kitchen. They have all the ingredients placed on the kitchen counter and are about to start. They read that the first thing to do is to break two eggs and mix them with the milk. The challenge is to figure out who will take care of the bowl and breaking the eggs, and who will pour the milk in a measuring cup? They want to prevent bumping into each other while they are reaching for the ingredients in the limited space available. By attending jointly to the ingredients on the kitchen counter, the two friends establish perceptual common ground (they share representations of objects and events) and become aware of each other's action opportunities and constraints (Curioni et al., 2017; Sebanz, Bekkering, & Knoblich, 2006). So, when the eggs are closest to a certain person, this person may decide to take on the task of breaking the eggs. When the other friend follows the other's gaze towards the eggs box, he or she would know that he or she needs to reach for the milk to start the joint action efficiently. In this scenario, representing what a person is attending to (through gaze cues) provides important information about his or her action goals. This information facilitates the selection of the appropriate actions to carry out the joint plan.

According to Brinck (2008), four different types of behavior are involved when referencing an object non-verbally. First, a *preparatory behavior* draws the observer's attention to the sender. Second, a *communicative-intent indicating behavior* signals the sender's attempt to share attention and interact with the observer. Third, a *referential behavior* orients the attention of the observer in the direction of the target object or event. Finally, an *essentially intentional behavior* orients back the attention of the observer and the sender to each other to make sure that they understand the situation.

Huang and Thomaz (2010) describe four different steps to reach joint attention. First, two agents need to pay attention to each other to become aware of and anticipate an upcoming interaction. This step matches with the first two types of behavior described by Brinck (2008). Second, one of the agents initiates joint attention by directing his or her attention to the object of interest. This agent also addresses the object using communicative channels like vocal comments. Then, the other agent responds to the request for joint attention by directing his or her attention to the referred object. These behaviors match with the referential behavior described by Brinck (2008). Third, to ensure that the observing agent attends to the object of interest, the initiating agent looks back and forth between the responding agent and the object. If necessary, the initiating agent performs additional actions to draw the attention of the responding agent to the object of interest. This step matches with the *essentially intentional behavior* described by Brinck (2008). Finally, the two agents establish joint attention while both attending to the object of interest and then continue with their interaction. It is important for the initiating agent to keep performing the third and fourth steps during the interaction in order to ensure that the joint attention is maintained.

According to Moore, Dunham, and Dunham (2014), joint attention has three, instead of four, phases. It begins with mutual gaze to establish attention, proceeding with referential gaze to direct the attention to the object of interest, and cycling back to mutual gaze to ensure that the experience is shared.

Based on a comparison of the steps described by Brinck (2008), Huang and Thomaz (2010), and Moore et al. (2014), there seem to be three major steps involved in the process of establishing joint attention. First, two agents need to initiate an interaction. Second, the initiating agent directs his or her attention to the object of interest with the aim to draw the attention of the responding agent to the object as well. Third, the initiating agent checks whether the responding agent is indeed attending to the referred object. If this is not the case, the second and third step are repeated until joint attention is established.

Joint Attention in Human-Robot Interaction

Efforts have been made to develop robots that can interpret human behavior. For example, Anjum, Ahmad, Rosa, Yin, and Bona (2014) developed an algorithm that can successfully recognize eight different complex human activities (like waving hello, looking at a watch, or picking something up from the ground) based on skeleton tracking, irrespective of the individual performing the activities and the position of the individual in front of the camera. However, when the range of recognizable human activities increases, differences in movement patterns will become smaller, which makes it harder to correctly classify similar behaviors. Another example of an effort is that Droeschel, Stückler, Holz, and Behnke (2011) implemented a camera system that enables a domestic service robot to perceive showing and pointing gestures of humans and match the estimated pointing directions and shown objects to objects in the robot's environment. This allows the robot to infer what the human is attending to, which in turn provides important information about likely future actions of this person. Because communication is bidirectional, humans also need to be able to interpret a robot's behavior and what it is attending to. One way is to make a robot establish joint attention.

For example, Yonezawa, Yamazoe, Utsumi, and Abe (2007) showed that a gaze-communicative stuffed-toy robot that establishes joint attention can effectively draw a user's attention to where the robot is looking, which results in perceptual common ground. Also, Sauppé and Mutlu (2014) implemented several deictic gestures (with contingent gazing behavior) on a humanlike robot to evaluate their communicative accuracy and perceived effectiveness in varying settings. Moreover, Huang and Thomaz (2010) showed that a humanoid robot that responds to joint attention (through gaze cues) initiated by a person, is perceived as more transparent, competent and socially interactive. In addition, the same study showed that a robot that ensures joint attention yields better performance in a human-robot interactive task and this behavior is perceived to be natural by humans.

Multiple studies showed that gaze cues of humanoid robots, when used congruently with speech references, enable effective joint attention by disambiguating spatial references in speech, improving task performance in joint action (Admoni, Datsikas, & Scassellati, 2014; Boucher et al., 2012; Mutlu, Huang, & Terrell, 2013). In addition, Mutlu et al. (2013) showed that robots that establish joint attention in a joint action, are perceived as being more competent. In a cooperative experiment by Boucher et al. (2012), a robot with articulated eyes and head provided verbal information specifying the target object that the participant had to grasp. Across three conditions, the information provided by gaze cues of the robot differed. In the first condition, there was no head or eye movement. In the second condition, the robot indicated the location of the target object by a coordinated eye and head movement (the onset of these cues preceded the speech command). In the third and final condition, the robot wore sunglasses, so only the movement of the head (but not the eyes) was visible. In the full gaze and sunglasses conditions, the mean reaction times and movement times of the participants were significantly shorter compared to the condition in which the head and eyes were fixed. These results illustrate that the robot's eye and/or head movements could be used by the participants to anticipate the specification of the location of the target object, which made the interaction more fluent.

Looking critically at the previous studies, the experimental task of anticipating which object the robot will refer to (before the onset of a speech cue) seems very simple compared to a real-world setting. First, the set of objects the robot could possibly refer to, was quite limited and clearly visible. Second, the link between the goal of the robot (indicate the object that the participant needs to grasp) and the subject of attention was very clear. In real-world settings, this link might be less clear. For example, going back to the example of two friends baking pancakes, one agent might look at the milk, but start reaching for a measuring cup (to separate the right amount of milk). In the experiments, the robot basically told the participants what to do (reach for a certain object), while in real life, people have to determine themselves how they should behave based on what they expect others will do. Finally, the range of likely behaviors that the robot could perform was very limited: verbally indicate a target object, but no movements that might interfere with participants' actions. In the baking example, both agents manipulate objects in their environment, which makes the joint action more complicated. Thus, it would be interesting to study whether the results of the previous studies also apply to more complicated settings.

Research Gap

Important to note is that the experiments in the studies in the previous section all involve (nearly) static robots that have a relatively high degree of human-likeness. This allows the use of gaze cues, which are intuitive and easy to understand by humans if implemented correctly. When using gaze cues in a joint action process, the following is important to remember: In a successful joint attention process, attention needs to be directed towards the agents themselves and then to an external object. Often, attention is only directed at an external object, which might not be sufficient: Admoni, Bank, Tan, Toneva, and Scassellati (2011) found that robot gaze does not reflexively cue human attention. Thus, when a robot is gazing at an object, this does not always mean that observers will notice the robot and its subject of attention. Natural gaze cues are difficult to implement in non-humanoid robots, so there is only little research about establishing joint attention in mobile non-humanoid robots. For example, in an interesting study conducted by Movellan and Watson (2002), it was found that ten-month old infants show sensitivity to a robot's gaze direction (indicated by the orientation of a square box), following the line of sight of a non-humanoid robot head. In addition, Szafir, Mutlu, and Fong (2015) explored different designs of small flying robots to visually express a robot's intent (flight directionality) through a LED ring attached to the robot. Participants were asked to predict as fast as possible whether the robot would fly to a specific target. Interestingly, one of the designs resembled human eye movements and this design improved participants' speed and accuracy predicting robot intent and their perceptions relating to the robot's usability the most. However, the *blinker* and *thruster* designs were also effective. In line with the GOADI theory, the researchers concluded that visual cues should convey high level aspects of flight intentions (the final destination of the robot) rather than low level corrections to flight paths. The two previously described studies suggest that robots do not necessarily have to look human in order to be able to make their intentions clear by directing an observer's attention to a target object. However, many non-humanoid robots do not allow displaying behavior that tries to reflect human gazing behavior. Therefore, the question remains: How should a non-humanoid robot (that does not have "eyes" or a "head" to simulate gaze cues) establish joint attention with a human agent to communicate its intentions?

Another area in which HRI research is lacking is industrial settings like warehousing environments. Such an environment is different from the isolated lab settings used in the aforementioned studies. For example, there are many distractions and noises because there are a lot of moving machines and people around. Verbal cues would not work well here (Bejerano, LeMasurier, & Yanco, 2018).

In addition, from the context and instructions given in the experiments described in the previous section, the participants were aware that they were going to interact with the robot in front of them. Therefore, the first step of establishing joint attention (in which two agents attend to each other to indicate that they want to interact with each other) was often not explicitly implemented. However, in most real-world scenarios, for example a distribution center where several order pickers interact with multiple robots, this first step is necessary to be able to continue with the process of establishing joint attention.

As non-humanoid robots are becoming increasingly utilized for collaborative tasks across many domains like industrial settings, it is important to investigate how the fluency of HRC in such a setting can be improved. Because research on joint attention in experimental HRC tasks has proven to be effective in enabling human partners to anticipate future behavior of a humanoid robot, it is interesting to study whether this can be applied to non-humanoid robots in an industrial environment as well.

Collaborative Case Picking at Vanderlande

An example of an industrial area in which mobile robots need to collaborate with humans is a warehousing environment. Vanderlande Industries is investigating the possibility of a use case where autonomous mobile robots cooperate with human order pickers in a collaborative case picking scenario. Currently, human case pickers work in these environments using forklifts to pick cases. However, this can be done more efficiently if people put the necessary products on top of autonomous mobile robots, that transport all products to the place where the final package is assembled. This saves the time needed to step up and down from the forklift and bringing the final order to a delivery point. In one of the concepts proposed by Vanderlande, multiple order pickers interact with multiple robots, instead of one robot following one order picker.

In manual order picking, order pickers receive instructions about their next picking activity through a headset or display that they carry with them. When adding robots, it should also be communicated on which robot a package should be placed. This is especially important in case picking. Because the cases are quite heavy, the time that order pickers spend lifting cases should be minimized. One option is to simply assign unique codes to each robot and instruct the order picker to put the picked package on the robot with a specific code, but this increases the mental workload of the order picker. In one of the concepts proposed by Vanderlande, a screen display will be attached to the robot that provides picking instructions to the human order pickers. This removes the need for order pickers to carry around a separate device (for example a headset or small display). In this scenario, the communication between the robots and the order pickers is of critical importance. The interaction between a human order picker and a mobile robot can be considered a joint action. In the collaborative case picking scenario, the joint goal is to move a selected package from one location to another. The robot has the information related to the order (like the locations of the cases that need to be collected) and the order picker is responsible for picking up and stacking the cases.

Research Aims

Research question. The aim of the current study is to research whether the positive effect of joint attention on the fluency of the HRC found in static humanoid robots also applies to mobile non-humanoid robots in a less experimental setting. Because research about HRI in industrial settings is largely missing, the context of the current study is a collaborative case picking scenario in a warehousing environment. Hence, the research question is as follows:

How does a mobile non-humanoid robot that establishes joint attention with a human partner affect the fluency of the human-robot interaction in a collaborative case picking scenario?

In the collaborative case picking scenario, there are two main objects that are relevant for the robot and a human order picker to pay attention to. First, because there are multiple robots and order pickers present in a warehousing environment, the robot and its partner need to pay attention to each other to become aware of the possibility of an interaction (this corresponds to the first step of a joint attention process). Second, there is a specific case that the robot and the human partner need to transport, so the robot needs to draw the attention of the human partner to the target object by performing a referential behavior (this corresponds to the second step of a joint attention process). Finally, to successfully establish joint attention, the robot needs to check whether the human order picker is indeed attending to the target object. If this is not the case, the previous step needs to be repeated.

Because the robots employed by Vanderlande are non-humanoid, gaze cues are difficult to implement. Therefore, an alternative method will be implemented: lights close to the cases will be used to indicate that the robot is attending to a specific case and the color of the lights attached to the robot will be used to indicate that the robot is attending to a specific person (each person in the warehousing environment will be assigned a unique color). Gaze cues do not allow a humanoid robot to direct attention to two different physical objects at the same time. However, light cues do allow this option. It is interesting to see whether this makes the use of joint attention more effective in improving the fluency of the HRC.

Hypotheses. As mentioned before, previous studies on collaboration tasks with humanoid robots and human partners indicated that robots that establish joint attention with their human partner lead to more fluent interactions (in terms of higher joint performance) and more positive perceptions of the robot. Because joint attention can be established through multiple cues (for example, gazing behavior, pointing behavior and speech references) and so far, no study exist that suggest that non-humanoid robots are not able to direct the attention of a human observer to a target object or event, it is expected that a non-humanoid robot can use joint attention to communicate its goals towards humans. This will make it easier for people to anticipate the future behavior of the robot and adjust their behavior appropriately. Hence, the main hypothesis is as follows:

H1: A robot that directs its attention to the target order picker and/or the target case leads to a more fluent collaboration and more positive perception of the robot than a robot that does not establish joint attention with its human partner.

The aim is to test two additional hypotheses. The two different behaviors used to pay attention to the target order picker and target case correspond to the first and second step of a joint attention process, respectively. In some of the previous studies on joint attention in HRI, the first step was not always implemented as the context already made clear that the participants were going to interact with the robot in front of them. However, in the collaborative case picking scenario, this might not be the case as multiple robots are supposed to interact with several different order pickers. Hence, the first step in which the robot pays attention to the target order picker is very important in order to continue with the next step. Directing attention to the target case informs an observer about the goal of the robot, but when the robot has not made clear that it wants to interact with a specific person, the observer might feel unsure about whether to assist the robot in reaching its goal. Therefore, the second hypothesis is as follows:

H2: A robot that directs its attention to the target order picker leads to a more fluent collaboration and more positive perception of the robot than a robot that directs its attention to the target case.

Because the two different behaviors used to pay attention to the target order picker and target case correspond to different phases of a joint attention process, it makes sense to expect that a robot that implements both phases in its behavior is superior to a robot that implements one phase. A robot that shows more communicate behavior makes it easier for a human partner to infer the goals and thereby, the future behavior of the robot. Therefore, the third and last hypothesis is as follows:

H3: A robot that directs its attention to both the target order picker and the target case leads to a more fluent collaboration and more positive perception of the robot than a robot that directs its attention to only one of these objects.

Approach. In order to test the hypotheses, an experiment was conducted at Vanderlande that recreates a collaborative case picking scenario in which a mobile robot may or may not have the intention to interact with the participant. The subject of the robot's attention differed across conditions. To analyze the effect of establishing joint attention, the joint performance (measured by the time needed by the participants to collect the target case and place it on the robot) and perceived fluency of the HRC were assessed. Because previous studies showed that a robot that establishes joint attention affects the perceptions of the robot, the perceived usefulness and ease of use of the robot were also assessed.

Method

Design

In the current study, a collaborative case picking scenario was recreated in which participants performed a joint action with a mobile non-humanoid robot. The joint goal was to move a specified case (indicated by a code representing the location of the case on a rack) from one location to another. The participants collected and placed the case on a pallet carried by the robot and the robot shared the location where the participant could find the case, and transported it. The subject of the robot's attention (while approaching the participant) was varied.

The experiment had a 2 (do or do not communicate the target order picker by directing attention towards it) by 2 (do or do not communicate the target case by directing attention towards it) factorial design. This resulted in four conditions related to the subject of the robot's attention: In the baseline condition, the robot did not communicate anything about the focus of attention. In the second condition, the robot directed attention to the target order picker by changing the color of the LED lights attached to the robot. Each participant was assigned a color (orange). In the third condition, the robot directed attention to the target case by turning on a (white) LED light close to the target case on the racks. In the fourth condition, the robot directed attention to both the target order picker and target case. In this situation, the color of the LED lights attached to the robot and the color of the LED light on the racks close to the target case matched.

The experiment had a within-subjects design, as all participants experienced each condition (in semi-random order) eight times. Within each condition, the robot passed the participant two out of eight times. The goal of these catch trials was to introduce uncertainty to the participants. This made anticipating the robot's behavior harder and thus, participants were more motivated to pay attention to the information communicated by the robot. The order of measurement and catch trials was randomized within each condition.

Participants

A total of thirteen participants (four women, nine men, $M_{age} = 34.46$ years, age range: 19-53 years) were sampled from the employee pool of Vanderlande. The selection criteria for participant recruitment were being able to come to the office of Vanderlande in Veghel (The Netherlands), being mobile (as some cases needed to be lifted), having a comprehension of the English language, and being unaware of the expected outcomes of the study (as the possibility existed that a colleague had attended a presentation about the current project). All participants were asked to bring corrective eye-wear if needed, resulting in normal or corrected to normal vision for all participants. To thank participants for their effort and time, they received a box of chocolates.

Power and sensitivity analyses. To make an estimate of the required sample size, an a priori power analysis was performed in G*Power (Faul, Erdfelder, Lang, & Buchner, 2007). For the main hypothesis, a one-tailed t-test that compares means to a constant was performed, aiming for an error rate of $\alpha = .05$ and power of $1 - \beta = .90$. To determine the effect size of interest, effect sizes found in similar studies were analyzed. The effect sizes for the movement time were much larger than the effect sizes for the perceived fluency of the HRC, so for this analysis, the second measure was evaluated. Hoffman and Breazeal (2007) found that anticipation in HRI improves the perceived fluency of the interaction with Cohen's $d = 1.015$. Based on the input values described above, the required sample size to test the first hypothesis turned out to be eleven participants.

Studies related to the secondary hypotheses were not found, which made performing a power analysis difficult. Instead, a sensitivity analysis for a study with eleven participants was conducted. It was expected that the effects described by the second and third hypotheses would have smaller effect sizes than Cohen's $d = 1.0$, because the conditions that are compared are more similar to each other (compared to the baseline condition). To test the second hypothesis, two conditions were compared with each other in a certain direction. Therefore, the statistical test in the sensitivity analysis was chosen to be a one-tailed paired t-test. Assuming a sample size of eleven participants and again aiming for an error rate of $\alpha = .05$ and power of $1 - \beta = .90$, an effect with a standardized size equal or greater than Cohen's $d_z = 0.95$ could be detected.

For the third and final hypothesis, the statistical test in the sensitivity analysis was chosen to be a one-tailed t-test that compares means with a constant. Assuming a sample size of eleven participants and aiming for an error rate of $\alpha = .05$ and power of $1 - \beta = .90$, an effect with a standardized size equal or greater than Cohen's $d_z = 0.95$ could be detected.

Although the sensitivity analyses indicated that effects with sizes smaller than Cohen's $d_z = 1.0$ could be detected, Cohen's $d_z = 0.95$ is not much smaller than Cohen's $d_z = 1.0$. The current project served as the starting point for research and the design of a collaborative mobile robot in a warehousing environment. When deciding to alter the design of a robot, the change should have a large effect on the fluency of the HRC, otherwise it is not worth the cost. Therefore, a minimal sample size of eleven participants was required, but the aim was to collect as many participants as feasible.

Measurements

In order to study the effect of a robot that establishes joint attention with a human order picker on their joint performance, the time was measured between the moment that the robot arrived at its destination and the moment that the participant had placed the target case on the pallet carried by the robot. These movement times were inferred from the video recordings made during the experiment. For each participant, six measurements of the movement time were collected per condition. In addition to these measurement trials, participants experienced catch trials. During these catch trials, the robot did not approach the participants, but it passed them. The goal of the catch trials was to introduce uncertainty to the participants. This made anticipating the robot's behavior harder and thus, participants were more motivated to pay attention to the information communicated by the robot. During catch trials, no movement time was measured.

To evaluate the subjective experience of the participants, a questionnaire was presented (see Appendix C). First, the perceived fluency of the HRC was assessed using survey questions based on the metrics of Paliga and Pollak (2021). Participants were asked to answer six questions on a 7-point Likert scale from 1 (*strongly disagree*) to 7 (*strongly agree*). More specifically, the selected questions were related to three different perspectives of subjective HRI, i.e., human-oriented, robot-oriented, and team-oriented. After finishing the data collection, a Cronbach's α scale reliability test was performed with all six items to create one variable that represents the perceived fluency of the HRC.

To assess the perceived usefulness and perceived ease of use of the robot, questions were adapted from the UTAUT model by Venkatesh, Morris, Davis, and Davis (2003), which is based on the work of Davis (1989). For example, the word *system* in the original questions was changed into *robot* and *job* into *collaborative case picking task*. Moreover, the original questions were phrased in such a way that it seems like participants are asked to imagine that they would use the system in a future task and base their evaluation on that. However, in the current study, the aim was to assess the robot based on the experience that participants just had. Therefore, the items were rephrased by removing the word *would*. As two of the original items related to the perceived ease of use did not fit with the current experiment, they were left out. One item actually measured two different aspects (*My interaction with the robot is clear and understandable*), so this item was split up into two separate items. In the end, six items were used to measure the perceived usefulness and five items to measure the perceived ease of use of the robot. These questions were also answered on a 7-point Likert scale. After finishing the data collection, two Cronbach's α scale reliability tests were performed to create one variable that represents the perceived usefulness of the robot and one variable that represents the perceived ease of use of the robot.

Finally, the last questionnaire (see Appendix D) contained some questions regarding the demographics of the participants (age, gender) and their previous experience with mobile robots and order picking.

Materials and Setting

To recreate the collaborative case picking scenario, several materials were used. They are described in the following subsections.

Racks with cases. In case picking scenarios, most cases are fairly heavy. As a heavy case requires more effort to lift it, people prefer to know where a heave case should be moved to before they pick it up. In that case, anticipating what an approaching robot is going to do is even more important, because a heavier object is more likely to get damaged when it drops. For the experiment, eighteen boxes of cat food were used. Each box weighed about two kilos and had a size of 39 cm x 28.5 cm x 19 cm. These cases were structured and labeled according to a three by six grid on two racks (see Figure 1). Both racks had a size of 182 cm x 136 cm x 52 cm. The height difference between two layers was about 50 cm. Each location on the racks was assigned a unique label, which was placed just below a case. These labels consisted of a number and a letter. The number (between one and six) indicated the column (starting from the right) and the letter (A, B or C) indicated the height (starting from the top).

LED lights. To enable the robot to draw attention to the target order picker and target case, twenty multi color LED lights were used (see Figure 2). These lights can emit light in thirteen different colors and the brightness levels can be adjusted. Each light can be controlled individually by a remote. By unscrewing the lights, the batteries can be replaced. Eighteen lights were attached to the racks, so that each location had a LED light placed next to its corresponding label (see Figure 1). The remaining two LED lights were attached to the robot (one above the instruction and one on the right front side, see Figure 3).

Sweat bands. By changing the color of the LED lights attached to the robot, the robot aimed to draw attention to the target order picker. Each order picker was assigned a unique color. To make clear to the participants which color they had been assigned, they were asked to wear sweat bands with a certain color (orange).

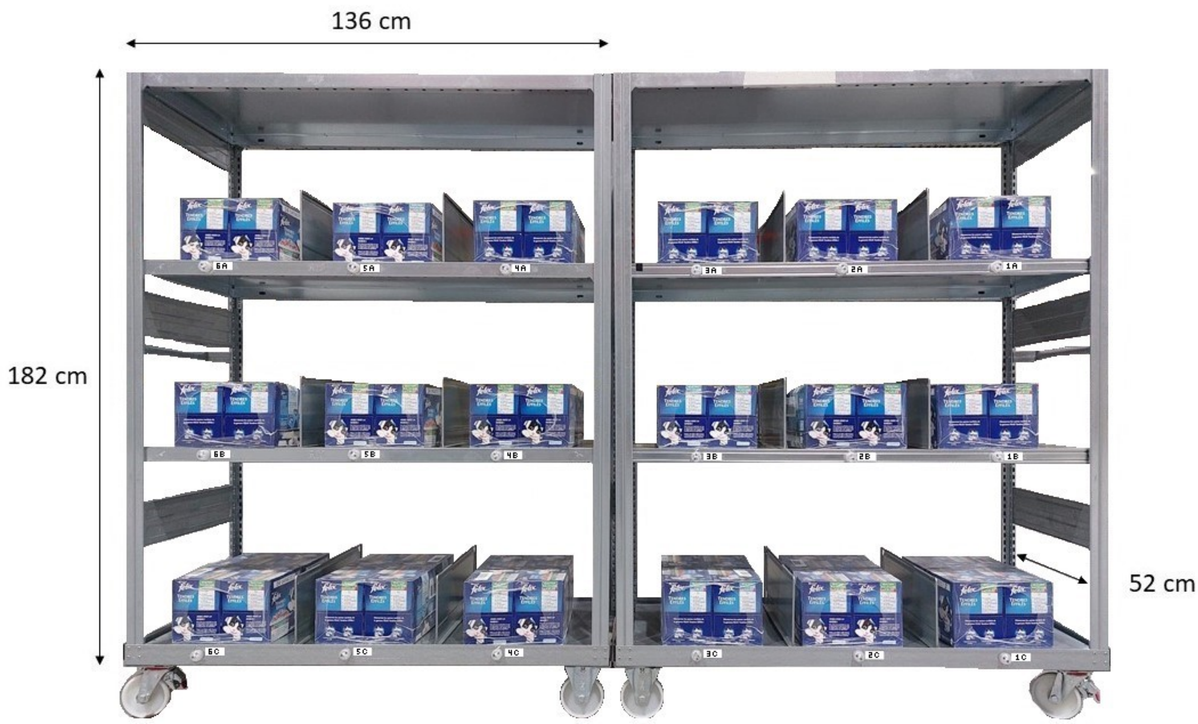

Figure 1. The eighteen cases used in the experiment were structured and labeled according to a three by six grid on two racks. LED lights were placed to the left of each label.

Forklift robot. In the experiment, participants interacted with the stacker automated electric (SAE) robot developed by Toyota[1] (see Figure 3). This robot can carry a pallet or load carrier on its forks. In the experiment, the robot carried a wooden pallet on its forks at a height of 55 cm above the ground. Its maximum speed is 2.2 m/s, but the speed dropped when it made a turn.

For safety reasons, the robot provided some signals using the red lights on the front and side (see Figure 3). First, the lights on either the left or right side of the robot blink to indicate that the robot is going to make a turn. Second, when the robot starts to move after it has been standing still for more than ten seconds, all red lights blink once. In addition, when the robot is moving, a blue light beam is pointing to the space in front of the robot (see Figure 3). This is implemented by robot's developer to make it easier for people working in aisles to become aware of the presence of the robot when

[1] https://toyota-forklifts.nl/automatisering-magazijn/geautomatiseerde-magazijntrucks

Figure 2. The LED lights used by the robot to draw attention to specific objects. The LED lights allow a range of colors and can be individually controlled by a remote.

the robot is blocked from their view. When the robot is standing still, the light beam disappears.

During the experiment, the robot was controlled by the experimenter through a laptop that had a remote desktop connection with a PC that had the necessary programs installed. The path of the robot consisted of a closed loop with a rectangle-like shape. On this path, four points were defined (see Figure 4). Point 1 was the starting position of the robot in each trial. This point was located between two big racks to limit the view participants had of the robot. Point 2 was located one meter in front of the first rack and point 3 was located one meter in front of the second rack. These points were at a distance of 1.2 meters from each other. Point 4 was just in front of the first point and it was used to make the robot pass the points in front of the racks (ordering the robot to move to point 1 when it was positioned at point 1 did not work). In the program an order was created to send the robot to the point set by the user (see Figure 4). Because the environment in which the experiment was conducted contained some obstacles, the program was used to draw lines at the corners of the obstacles to make sure that the robot would not collide.

In all conditions, the robot indicated the location of the target case through a display. Otherwise it would have been very difficult for participants to perform their picking tasks in the baseline condition. This is different from the manipulation of the LED lights on the racks, because presenting the code does not directly draw the

Figure 3. The SAE robot used in the experiment (developed by Toyota) and its communicative signals. Blinking safety lights communicated the navigational intent of the robot (start and stop moving and making a turn). Two LED lights (emitting orange light in the figure) were manipulated during the experiment to enable the robot to indicate the target order picker. A piece of paper (attached with clips) was used to present a code to the participants that matched with one of the locations on the racks. Because of the safety measures implemented by the robot's developer, a blue spot was projected on the floor in front of the robot. This makes it easier for people working in aisles to become aware of the presence of the robot when the robot is blocked from their view.

attention of the participants to the physical object. Instead of a screen display, pieces of paper were attached one at a time to the right side of the robot with a clamp (see Figure 3).

Video recordings. To measure the movement times of the participants, video recordings (25 frames per second) were made of their interaction with the robot. At the location where the experiment took place, several CCTV camera's were present. The most appropriate camera was used. It was important that this camera was able to capture the moment when the robot came to a halt in front of the racks and the moment that a participant placed a case on the pallet carried by the robot. To play

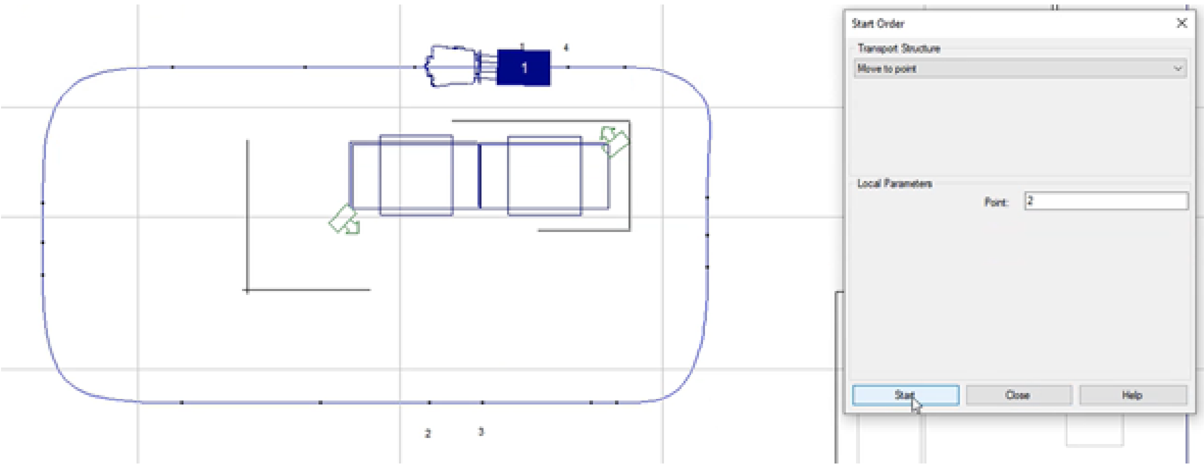

Figure 4. The program used to control the SAE robot by sending it to one of the four points.

back the video recordings, the software XProtect Smart Client[2] was used. This software allowed to analyze each single frame with the corresponding timestamp.

Setup. The experiment was conducted in the Advanded Design Center of Vanderlande located in Veghel, The Netherlands. As there were multiple projects with different machines going on at the same time, the level of background noise was quite high (which is similar to real-world order picking settings). For safety reasons, participants wore safety shoes and people with long hair tied their hair into a ponytail. Figure 5 shows an overview of the setup of the experiment. At the start of a trial, the participants were unable to view the robot as there were two obstacles between them and the participants were instructed to face the table. The orange bar on the robot's path represents a piece of tape on the floor used by the experimenter to determine when the participants should start walking towards the racks (when the front of the robot was touching the line). This was done to make the setup more similar to a collaborative case picking scenario. Usually, order pickers perform picking times quickly after another, so they do not have very much time in-between instructions and their mental workload is high. If the participants were allowed to view the robot directly when the robot started moving, participants would have at least 30 seconds to focus solely on the behavior of the robot. This is not realistic and therefore, the time available to

[2] https://www.milestonesys.com/

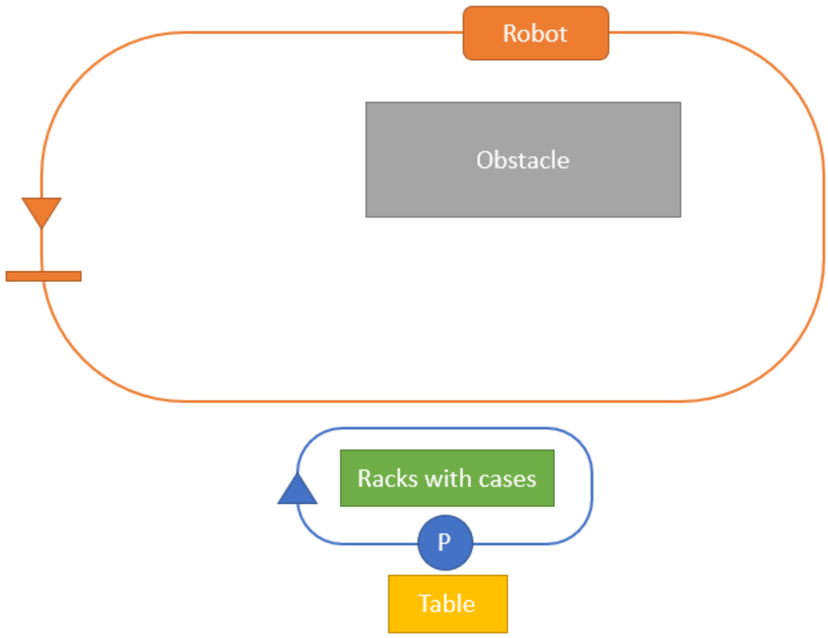

Figure 5. An overview of the setup of the experiment. The orange line indicates the robot's path and the blue line indicates the participant's path.

participants to observe the robot was limited. The participants used the table to fill out the subjective experiences surveys on paper.

Procedure

Table 1 shows an overview of the conditions, number and type of trials, and surveys the participants experienced during the experiment. Before the start of the experiment, participants were asked to read and sign an informed consent form (see Appendix A). If participants gave consent, they were asked to put on safety shoes, tie their hair into a ponytail (if they had long hair), and sign a document that stated that participants should behave appropriately around the robot. Because most participants did not have any experience with being in close proximity with a robot bigger than themselves, participants were asked to step in front of the robot while it was moving. By demonstrating that the robot would stop in time, the participants were assured that they would not get hit by the robot. Next, the participants received instructions for the experiment (see Appendix B). To make sure that the participants understood their task, three practice trials were presented (including one catch trial) according to the baseline condition.

Table 1

Overview of conditions, trials and surveys that make up one experiment

Phase	Condition	Number of measurement trials	Number of catch trials	Order of trials
Practice (no actual measurements)	Baseline (fixed)	2	1	The same for each participant
Experiment	Baseline (fixed)	6	2	Randomized for each participant
	Subjective experience survey			
	Order picker (randomized)	6	2	Randomized for each participant
	Subjective experience survey			
	Order picker & case (randomized)	6	2	Randomized for each participant
	Subjective experience survey			
	Case (randomized)	6	2	Randomized for each participant
	Subjective experience survey			
End	Demographics survey			

At the start of a trial, participants were standing at a table and their view of the robot was blocked (see Figure 5). This allowed the experimenter to prepare the robot for each trial without influencing the information available to the participants before the start of the trial. When the robot passed the orange bar in Figure 5, the experimenter said "Start" and the participants walked to the other side of the racks. The participants always arrived earlier than the robot. On the right side of the robot, a code was presented that indicated the location of the target case (see Figure 3). This location was randomly determined beforehand.

In a measurement trial, the participant matched with the target order picker of the robot, so the robot stopped in front of one of the racks. Next, participants picked the target case and put it on the pallet carried by the robot. When the experimenter had made sure that the case was placed on the robot, the experimenter instructed the robot from a distance to move along and eventually back to the starting point behind the obstacle. At the same time, the participants walked back to the table, so that the experimenter could prepare the next trial.

In the case of a catch trial, the participant did not match with the target order picker of the robot, so the robot passed the participant and the target case remained at the shelve. Before the start of the experiment, the participants were informed that the robot might pass them. The explanation was that in a real warehousing environment, multiple robots and order pickers are present, so participants should not assume that every robot they see has the intention to interact with them.

No measurements were taken during the three practice trials. Next, participants were presented with eight trials according to the baseline condition. Two out of eight trials were catch trials. To create uncertainty, the order of the trials was randomized. For each measurement trial, the movement time of the participant was measured by watching the recordings of the experiment later on. After a block of trials, participants were asked to fill out a questionnaire regarding their subjective experience with the robot.

Then, participants were presented with three more blocks of eight trials (including two catch trials) and questionnaires. The behavior of the robot differed between the blocks based on the condition. In one condition, the robot directed attention to the target order picker. This was done as follows: The participants were assigned a color before the start of the experiment (orange). If the robot intended to interact with the participant, the LED lights attached to the robot emitted light of the same color (orange). If the robot did not have the intention to interact with the participant, the LED lights attached to the robot emitted light of a different color than orange (blue). In another condition, the robot directed attention to the target case. This was done by turning on a LED light (with a neutral white color) next to the label associated with the target case. In the final condition, the robot directed attention to both the target order picker and the target case. Again, LED lights attached to the robot and close to the target case were used, but now the colors of both lights matched the color assigned to the participant (orange). Although all participants experienced the baseline condition first, the order in which the other three conditions were experienced was randomized among participants.

At the end of the experiment, participants filled out a questionnaire regarding demographics and their previous experience with mobile robots and order picking. Finally, participants took off the safety shoes, were debriefed and thanked for their participation.

In total, the participants completed three practice trials and 32 normal trials, of which eight trials were catch trials. Each trial took about 105 seconds to complete and the questionnaire (that was administered four times) required about 2 minutes. Adding the time required for giving instructions and debriefing, one experiment lasted about 90 minutes.

Statistical Analysis

All statistical analyses were done using STATA 17[3].

Pre-processing. Before starting with the statistical tests, the data was pre-processed. First, the movement times were inferred from the video recordings. To accurately determine when the robot stopped moving and the participants had placed a case on the pallet carried by the robot, the videos were played back frame by frame using the software XProtect Smart Client. The movement time for each trial was

[3] https://www.stata.com/

determined by calculating the difference between the timestamp at which the participant had placed a case on the pallet carried by the robot (and their hands were about to let go of the case) and the timestamp at which the robot stopped moving to interact with the participant (this was indicated by the disappearance of the blue spot on the floor). For each participant, there were six measurements of the movement time per condition. After having checked for outliers in the individual measurements, an average was calculated of the movement time per condition per participant.

Second, the reliability of the scale items measuring the perceived fluency of the HRC, the perceived usefulness and perceived ease of use of the robot was assessed using Cronbach's α. If removing an item increased the value for Cronbach's α, it was not included when creating the final variables (by averaging the scores of the separate items).

Finally, the data were checked for missing values and outliers. To check for outliers, several methods were implemented. First, a graph was created for each of the four dependent variables in which the data points of each participant were plotted per condition. Data points belonging to the same participant were connected with a line and assigned the same color. This provided insights about the patterns in the data and whether some data points deviated from these patterns. Second, the values of the dependent variables were standardized within each condition. Because the data was not normally distributed in every condition and the sample size relatively small, a z-score was considered a potential outlier if its absolute value was equal or greater than the following criteria: $(n-1)/\sqrt{n}$ (where n is the sample size). Second, data values that were outside of the 75th percentile + 1.5 $\times$ inter quartile range (IQR) or the 25th percentile - 1.5 $\times$ IQR, were considered potential outliers. Finally, the data were checked for extreme outliers, which entail the data values that were outside of the 75th percentile + 4 $\times$ IQR or the 25th percentile - 4 $\times$ IQR.

Confirmatory analyses. To test the main hypothesis, contrast analyses were performed according to guidelines provided by Haans (2018). To investigate whether the mean movement time in the conditions in which the robot directed attention to the target order picker, target case, or both was lower than in the baseline condition, the contrast weights were set to 1 for the baseline condition and -1/3 for each of the other conditions. To investigate whether the mean scores on perceived fluency of the HRC, the perceived usefulness and ease of use of the robot were higher in the conditions in which the robot directed attention to the target order picker, target case, or both, than in the baseline condition, the contrast weights were set to -1 for the baseline condition and 1/3 for each of the other conditions.

To test the second hypothesis, one-tailed paired t-tests were performed to compare the condition in which the robot directed attention to the target order picker with the condition in which the robot directed attention to the target case, in terms of the average movement time, perceived fluency of the HRC, the perceived usefulness and ease of use of the robot.

To test the third hypothesis, several contrast analyses were performed. To investigate whether the mean movement time in the condition in which the robot directed attention to both the target order picker and target case, was lower than in the conditions in which the robot directed attention to only one of these objects, the contrast weights were set to -1 for the full joint attention condition and 1/2 for the other two conditions. To investigate whether the mean scores on perceived fluency of the HRC, the perceived usefulness and ease of use of the robot were higher in the condition in which the robot directed attention to both the target order picker and target case than in the conditions in which the robot directed attention to only one of these objects, the contrast weights were set to 1 for the full joint attention condition and -1/2 for the other two conditions.

Assumptions. Before testing the hypotheses, several assumptions related to the statistical tests were checked. For contrast analysis, the assumptions are similar to those of conventional ANOVA: The data should be normally distributed in each condition and the variances of the differences between all possible pairs of within-subject conditions should be equal (sphericity assumption). However, this second assumption does not apply to contrast analysis as violations of sphericity cannot occur with focused tests (Haans, 2018). The main assumption of a paired t-test is that the differences between paired data points are approximately normally distributed.

Results

Checking Assumptions

Before testing the hypotheses, the relevant assumptions (described in the Method section) were checked.

Outliers. First, the data were checked for the presence of outliers using several methods (like the line graphs in Appendix E). The scores for the perceived usefulness of the robot contained two outliers. As can be seen in Figure E3, the data points of participant 9 in the *order picker* and *order picker & case* conditions deviated from the general pattern. Moreover, these points were marked as (extreme) outliers based on the IQR method. A possible reason could be that the participant accidentally thought that the response scale was from seven to one instead of one to seven, because if the scores would be flipped, they would have fit with the general pattern. However, this was unlikely as this participant did not show this possible mistake in the scores for the perceived fluency of the HRC and the perceived usefulness of the robot. Because of possible issues related to the normal distribution of the data, it was decided to exclude the data of the ninth participant from further analyses.

Normality. The data related to the four dependent variables were normally distributed within all conditions, except for the average movement time data and the perceived ease of use scores in the *order picker & case* condition. However, a repeated measures ANOVA is considered fairly robust against violations of this assumption. Interesting to note is the presence of a ceiling effect in the subjective experience scores in the *order picker & case* condition. There was little variance in the scores, because most responses were around six or seven (on a scale from one to seven). This was probably the result of participants giving relatively high scores in the baseline condition, which meant that there was not much room to give a much higher score in the *order picker & case* condition.

The differences between paired data points from the condition in which the robot directed attention to the target order picker and the condition in which the robot directed attention to the target case, were normally distributed for all dependent variables.

Sphericity. The sphericity assumption was met for the average movement time data and the perceived fluency of the HRC scores, but not for the perceived usefulness (p = .001) and perceived ease of use (p = .012) scores. To correct for this violation, the Greenhouse-Geisser correction was used.

Descriptive Analysis

Table 2 depicts the descriptive statistics of the variables used in the data analyses. The variables *perceived fluency of the HRC*, *perceived usefulness of the robot*, *perceived ease of use of the robot*, were created by averaging the scores on the corresponding items that were used to measure these dependent variables. A Cronbach's *alpha* scale reliability test showed that the variable representing the perceived fluency of the HRC

Table 2

Descriptive analysis of all variables

Variable	*M*	*SD*	Min	Max	Notes
Age	35.75	9.96	22	53	
Gender	0.25	.44	0	1	Male = 0, Female = 1
Average movement time (ms)	2202.44	1623.40	12	6934	
Perceived fluency of the HRC	5.78	1.01	3	7	
Perceived usefulness of the robot	5.76	1.13	2.67	7	
Perceived ease of use of the robot	5.90	1.15	2	7	

had an α value of $\alpha = .89$. This value could not be further increased by excluding one of the six items. The Cronbach's α for the variable representing the perceived usefulness of the robot was $\alpha = .96$. Although this value could be slightly increased by excluding two of the six items, all items were included because the value wass already very high. The Cronbach's α for the variable representing the perceived ease of use of the robot was $\alpha = .93$. Again, this value could be slightly increased by excluding one of the five items, but all items were included, because the gain was negligible.

Confirmatory Analyses

The current section describes the results of the statistical tests that were performed to test the hypotheses.

Effect of joint attention on fluency of HRC and perception of robot compared to baseline. The first hypothesis described the assumption that a robot that directs its attention to the target order picker and/or the target case leads to a more fluent collaboration and more positive perception of the robot than a robot that does not establish joint attention with its human partner.

First, the average movement was analyzed as a measure of the fluency of the HRC. Figure 6 shows the mean average movement time for each condition. The average movement time was longest in the baseline condition, shorter when the robot directed attention to the target case, even shorter when the robot directed attention to the target order picker, and shortest when the robot directed attention to both objects of interest. The results of the contrast analysis testing whether the average movement time was shorter in the conditions in which the robot directed attention to at least one of the objects of interest compared to the baseline condition, indicated that the first hypothesis could explain $\eta^2{}_{alerting} = 68.59\%$ of the observed differences between the baseline condition and the other three conditions ($F(1{,}11) = 25.12$, $p < .001$) with a contrast of $C = 1642.64$ ms.

Second, in addition to the observed fluency of the HRC, the perceived fluency of the HRC was analyzed. Figure 7 shows the mean scores for the perceived fluency of the HRC for each condition. This figure indicates that the HRC was perceived to be slightly less fluent when the robot indicated the target case compared to the baseline condition, which is not in line with the first hypothesis. On the contrary, the HRC was perceived

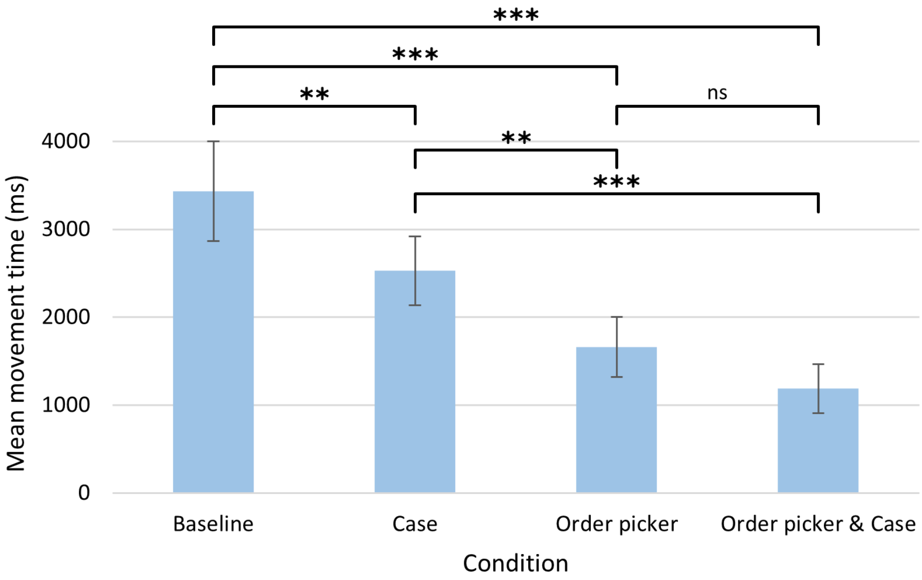

Figure 6. The mean average movement time (ms) for each condition. Error bars represent standard errors. The symbol * indicates the significance level of the corresponding Tukey's HSD post-hoc test (ns $p \geq .05$, * $p < .05$, ** $p < .01$, *** $p < .001$).

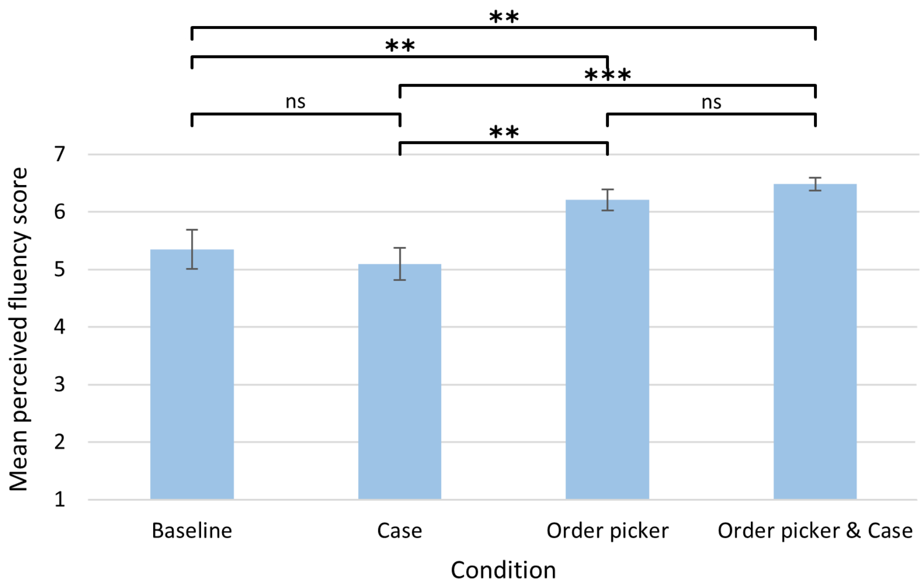

Figure 7. The mean perceived fluency of the HRC scores for each condition. Error bars represent standard errors. The symbol * indicates the significance level of the corresponding Tukey's HSD post-hoc test (ns $p \geq .05$, * $p < .05$, ** $p < .01$, *** $p < .001$).

to be more fluent when the robot indicated the target order picker, but adding information about the target case did not improve the perceived fluency of the HRC much further. The results of the contrast analysis testing whether the perceived fluency was higher in the conditions in which the robot directed attention to at least one of the objects of interest compared to the baseline condition, indicated that the first hypothesis could explain ${\eta^2}_{squared} = 19.11\%$ of the observed differences between the baseline condition and the other three conditions ($F(1{,}11) = 6.02$, $p = .032$) with a contrast of $C = 0.58$ scale points on the 7-point Likert scale.

Third, the perceived usefulness of the robot was analyzed. Figure 8 shows the mean scores for the perceived usefulness of the robot for each condition. This figure shows that the robot that directed attention to the target case was perceived as slightly more useful than the robot in the baseline condition. The robot that indicated the target order picker was perceived as even more useful and the robot that indicated both objects of interest was perceived as being most useful. The results of the contrast analysis testing whether the perceived usefulness was higher in the conditions in which the robot directed attention to at least one of the objects of interest compared to the baseline condition, indicated that the first hypothesis could explain ${\eta^2}_{squared} = 47.04\%$ of the observed differences between the baseline condition and the other three conditions ($F(1{,}11) = 18.03$, $p = .001$) with a contrast of $C = 0.89$ scale points on the 7-point Likert scale.

Fourth, the perceived ease of use of the robot was analyzed. Figure 9 shows the mean scores for the perceived ease of use of the robot for each condition. This figure indicates that the robot that directed attention to the target cased was perceived as about equally easy to use as the robot in the baseline condition. This is not in line with the first hypothesis. The robot that indicated the target order picker was perceived as easier to use, but adding information about the target case did not improve the perceived easy of use of the robot much further. The results of the contrast analysis testing whether the perceived ease of use was higher in the conditions in which the robot directed attention to at least one of the objects of interest compared to the

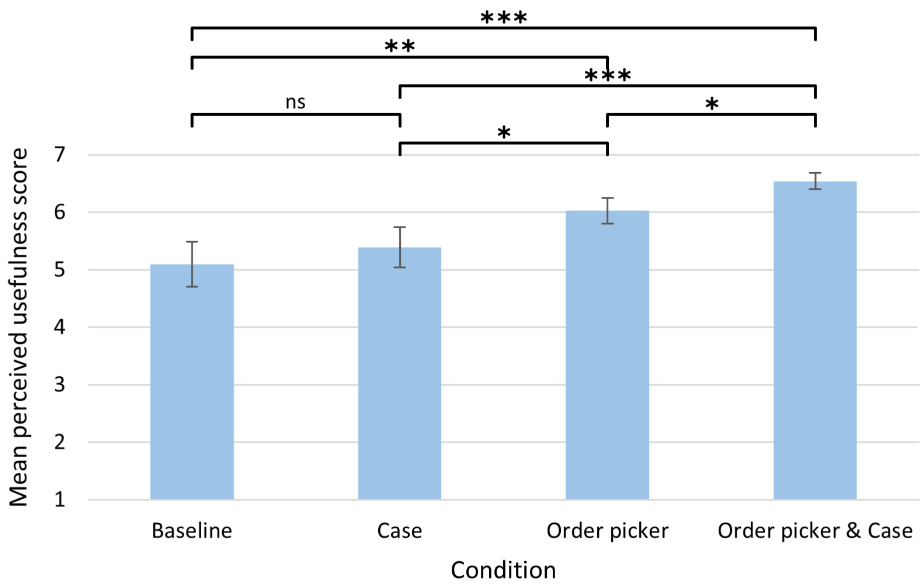

Figure 8. The mean perceived usefulness of the robot scores for each condition. Error bars represent standard errors. The symbol * indicates the significance level of the corresponding Tukey's HSD post-hoc test (ns $p \geq .05$, * $p < .05$, ** $p < .01$, *** $p < .001$).

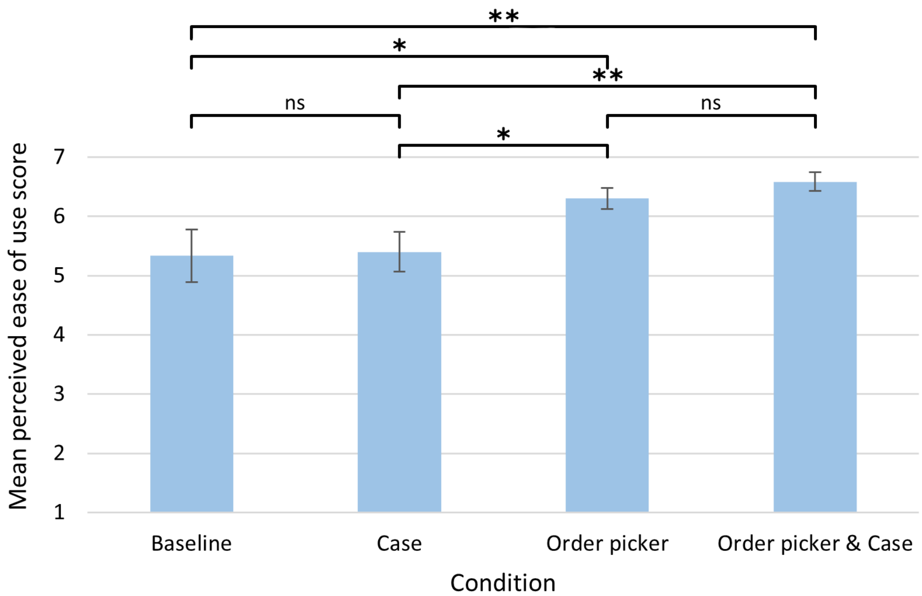

Figure 9. The mean perceived ease of use of the robot scores for each condition. Error bars represent standard errors. The symbol * indicates the significance level of the corresponding Tukey's HSD post-hoc test (ns $p \geq .05$, * $p < .05$, ** $p < .01$, *** $p < .001$).

baseline condition, indicated that the first hypothesis could explain $\eta^2{}_{squared} = 36.27\%$ of the observed differences between the baseline condition and the other three conditions (F(1,11) = 5.71, p = .036) with a contrast of C = 0.76 scale points on the 7-point Likert scale.

An overview of the results of the analyses performed to test the first hypothesis, is shown in Table 3.

Table 3

Results of contrast analyses testing hypothesis 1

Variable	C	df	F	p	$\eta^2{}_{alerting}$
Average movement time (ms)	1642.64	1	25.12	$<$ .001	68.59%
Perceived fluency of the HRC	0.58	1	6.02	.032	19.11%
Perceived usefulness of the robot	0.89	1	18.03	.001	47.04%
Perceived ease of use of the robot	0.76	1	5.71	.036	36.27%

Effect of different steps involved in establishing joint attention on fluency of HRC and perception of robot. The second hypothesis stated that a robot that directs its attention to the target order picker leads to a more fluent collaboration and more positive perception of the robot than a robot that directs its attention to the target case.

First, Figure 6 indicates that the average movement time was shorter in the *Order picker* condition than in the *Case* condition. The results of a one-tailed paired t-test confirmed that participants were significantly faster when the robot indicated the order picker it wanted to interact with (M = 1658.58, SD = 1183.17) compared to when the robot directed attention to the target case (M = 2529.25, SD = 1358.60); t(11) = 3.69, p = .002, 95% CI [351.19, 1390.15], Cohen's d_z = 1.07.

Second, Figure 7 indicates that the HRC in the *Order picker* condition was perceived as more fluent than the HRC in the *Case* condition. The results of a one-tailed paired t-test confirmed that participants perceived the collaboration with the robot that indicated the order picker it wanted to interact with as more fluent (M =

6.21, SD = .64) than the one with the robot that directed attention to the target case (M = 5.10, SD = .96); $t(11) = 5.02$, $p < .001$, 95% CI [.62, 1.60], Cohen's $d_z = 1.45$.

Third, Figure 8 indicates that the robot in the *Order picker* condition was perceived as being more useful than the robot in the *Case* condition. The results of a one-tailed paired t-test confirmed that participants perceived the robot that indicated the order picker it wanted to interact with as more useful (M = 6.30, SD = 0.77) than the robot that directed attention to the target case (M = 5.39, SD = 1.20); $t(11)$ = 3.75, p = .002, 95% CI [0.26, 1.01], Cohen's d_z = 1.08.

Fourth, Figure 9 indicates that the robot in the *Order picker* condition was perceived as easier to use than the robot in the *Case* condition. The results of a one-tailed paired t-test confirmed that participants perceived the robot that indicated the order picker it wanted to interact with as easier to use (M = 6.30, SD = .62) than the robot that directed attention to the target case (M = 5.40, SD = 1.16); $t(11)$ = 3.29, p = .004, 95% CI [0.30, 1.50], Cohen's d_z = 0.95.

Table 4 shows an overview of the results of the analyses performed to test the second hypothesis.

Table 4

Results of one-tailed paired t-tests testing hypothesis 2

Variable	*df*	*t*	*p*	95% CI	Cohen's d_z
Average movement time (ms)	11	3.69	.002	[351.19, 1390.15]	1.07
Perceived fluency of the HRC	11	5.02	< .001	[0.62, 1.60]	1.45
Perceived usefulness of the robot	11	3.75	.002	[0.26, 1.01]	1.08
Perceived ease of use of the robot	11	3.29	.004	[0.30, 1.50]	0.95

Note. Cohen's d_z refers to the standardized mean difference effect size for within-subjects designs.

Effect of combining different steps involved in establishing joint attention on fluency of HRC and perception of robot. The third hypothesis stated that a robot that directs its attention to both the target order picker and the target case leads to a more fluent collaboration and more positive perception of the robot than a robot that directs its attention to only one of these objects.

First, the average movement was analyzed as a measure of the fluency of the HRC. Figure 6 shows that the average movement time was shortest in the *Order picker Case*. The results of the contrast analysis testing whether the average movement time was shorter in the conditions in which the robot directed attention to both the target order picker and target case compared to the conditions in which the robot directed attention to only one of these, indicated that the third hypothesis could explain $\eta^2{}_{alerting} = 18.56\%$ of the observed differences ($F(1,11) = 9.59$, $p = .010$) with a contrast of $C = 906.42$ ms.

Second, in addition to the observed fluency of the HRC, the perceived fluency of the HRC was analyzed. Figure 7 indicates that the HRC was perceived to more fluent in the *Order picker Case* condition compared to the *Case* condition, but equally fluent in comparison to the *Order picker* condition. The results of the contrast analysis testing whether the perceived fluency of the HRC was higher in the conditions in which the robot directed attention to both the target order picker and target case compared to the conditions in which the robot directed attention to only one of these, indicated that the third hypothesis could explain $\eta^2{}_{squared} = 34.67\%$ of the observed differences ($F(1,11) = 13.07$, $p = .004$) with a contrast of $C = 0.83$ scale points on the 7-point Likert scale.

Third, the perceived usefulness of the robot was analyzed. Figure 8 shows that the robot in the *Order picker Case* condition was perceived as most useful. The results of the contrast analysis testing whether the robot was perceived as more useful when it directed attention to both the target order picker and target case compared to when it directed attention to only one of these, indicated that the third hypothesis could explain $\eta^2{}_{squared} = 36.75\%$ of the observed differences ($F(1,11) = 10.26$, $p = .008$) with a contrast of $C = 0.83$ scale points on the 7-point Likert scale.

Fourth, the perceived ease of use of the robot was analyzed. Figure 9 indicates that the robot was perceived as easier to use in the *Order picker Case* condition compared to the *Case* condition, but about equally easy to use in comparison to the *Order picker* condition. The results of the contrast analysis testing whether the robot was perceived as easier to use when it directed attention to both the target order picker and target case compared to when it directed attention to only one of these, indicated that the third hypothesis could explain $\eta^2{}_{squared} = 29.93\%$ of the observed differences ($F(1{,}11) = 17.37$, $p = .002$) with a contrast of $C = 0.73$ scale points on the 7-point Likert scale.

An overview of the results of the analyses performed to test the third hypothesis, is shown in Table 5.

Table 5

Results of contrast analyses testing hypothesis 3

Variable	C	df	F	p	$\eta^2{}_{alerting}$
Average movement time (ms)	906.42	1	9.59	.010	18.56%
Perceived fluency of the HRC	0.83	1	13.07	.004	34.67%
Perceived usefulness of the robot	0.83	1	10.26	.008	36.75%
Perceived ease of use of the robot	0.73	1	17.37	.002	29.93%

Exploratory Analyses

In addition to testing the hypotheses, exploratory analyses were performed. They are discussed in the following sections.

Role of experience on fluency of HRC and perceptions of robot. As a control, the demographics questionnaire contained an open question asking participants about their experience with order picking prior to the experiment. Only participant 1 had some experience with order picking from the past. Therefore, not enough data was collected to say something meaningful about the effect of experience with order picking on the movement time or perceptions of the robot. Looking at the line of participant 1 in the line graphs in Appendix E, the data did not suggest that the differences between conditions were different for participants with prior experience.

The demographics questionnaire also contained an open question asking participants about their experience with mobile robots prior to the experiment. Only participant 10 had some experience with mobile robots, but these robots were different (smaller) from the robot used in the current study. Therefore, not enough data was collected to say something meaningful about the effect of experience with mobile robots on the movement time or perceptions of the robot.

Correlations among variables. To investigate the relationships between the different dependent variables, a correlation matrix was analyzed for each condition. As can be seen in Table F1, the scores for the perceived fluency of the HRC and the perceived usefulness of the robot correlated significantly in the *Case* condition ($r = .77$, $p = .020$) and the *Order picker* condition ($r = .73$, $p = .041$). In addition, the scores for the perceived usefulness of the robot and the ease of use of the robot correlated significantly ($r = .89$, $p < .001$) in the *Order picker* condition.

One-on-one comparisons among conditions. Because the hypotheses could not explain all of the variance in the observed data, further exploratory analyses were performed in which each condition was compared with the other conditions in terms of the four dependent variables. For this purpose, two-way repeated measures ANOVA's were performed (see Table F2), followed up by Tukey's honestly significant difference (HSD) post-hoc tests. The results regarding the significance levels are visualized in Figures 6-9. Additional statistics can be found in Tables F3-F6. Overall, most conditions differed significantly from each other. However, when the robot directed attention to the target case, participants' perceptions of the fluency of the HRC and the robot did not significantly differ from the baseline condition, but their movement times did. In addition, the movement time, perceived fluency of the HRC and the perceived ease of use of the robot did not change when the robot directed attention to the target case in addition to the target order picker, but the perceived usefulness of the robot did.

Discussion

As robots need to collaborate with humans in industrial settings, it is important for the fluency of the interaction that humans can anticipate the behavior of their robotic partner. Communicating the goals of the robot aids this anticipation. Studies on humanoid robots showed that joint attention allows inferring the goals of such a robot. In the current study, an experiment was conducted to study the effect of a non-humanoid robot that establishes joint attention on the joint performance and the perceptions of the robot in a collaborative case picking scenario. While a stacker robot approached a participant, the subject of the robot's attention was varied across conditions (baseline, target case, target order picker, or both), which was indicated by LED lights.

The main hypothesis, stating that a robot that directs its attention to the target order picker and/or the target case leads to a more fluent collaboration and more positive perception of the robot than a robot that does not establish joint attention with its human partner, could explain a significant part of the differences between the baseline condition and the conditions with joint attention. The one-on-one comparisons in the exploratory analysis confirmed that the joint performance (in terms of the average movement time of the participant) was significantly higher in all conditions with joint attention compared to the baseline condition. The finding of a significant difference between the baseline condition and the condition in which the robot directed attention towards the target case (with Cohen's $d_z = 1.28$), was in line with the findings of Boucher et al. (2012), who found an significant effect on movement time with a size of Cohen's $d_z = 2.84$.

However, the perceptions of the participants did not differ significantly between the baseline condition and the condition in which the robot directed attention only towards the target case. So, although the actual fluency of the interaction improved when the robot indicated the case it aimed to collect (in terms of a decrease in the average movement time), participants did not perceive the collaboration as more fluent (Cohen's $d_z = 0.36$). This is in contrast to the findings of Hoffman and Breazeal (2007), who found that a robot that anticipates human behavior in a simulated factory setting improves the perceived fluency of the interaction with Cohen's $d = 1.015$. However, the agent that anticipates was different in this study. A correlation matrix indicated that the average movement time did not significantly correlate with the scores for the perceived fluency of the HRC in all four conditions. A possible reason for this is that indicating only the target case did not enable order pickers to anticipate whether the robot intended to interact with them (the first step of establishing joint attention was skipped), so they felt unsure about the robot's intentions. But when the robot stopped to interact with them, they had already found and moved to the target case, which decreased the time needed to locate and pick up the case. In addition, an external display attached to the right side of the robot indicated the location of the target case in each condition. So, participants did not necessarily need to infer the target case from the light cues on the racks (in the *Case* and *Order picker & Case* conditions). However, this information was available earlier and easier to process. One LED light was likely to pop out (bottom-up processing), which automatically drew the attention of the participants to the right case. When only a written code was available, participants still needed to search for the right location out of eighteen possible locations (top-down processing). Because the labels were rather simple (one number and a letter), it was easy for the participants (especially after some repetitions), to locate the target case quickly. Initially, the idea was to use more complicated labels, but a participant in a pilot study indicated that it was too complex for people without any experience in order picking. As real order pickers have practiced a lot with searching for a specific location based on a code, complex labels are not experienced as such after a while. Therefore, it was decided to use simple labels in the experiment. But if the experiment had been performed with more complex labels, the effect of communicating the target case through LED lights would probably have been larger (and significant).

The second hypothesis stated that a robot that directs its attention to the target order picker leads to a more fluent collaboration and more positive perception of the robot than a robot that directs its attention to the target case. This theory was supported by the data, as the average movement time was smaller, and the perceived fluency of the HRC, the perceived usefulness and ease of use of the robot were higher when the robot performed the first step (in contrast to only the second step) of establishing joint attention. These results confirm that the first two steps of establishing joint attention have indeed different functions.

The third and final hypothesis, stating that a robot that directs its attention to both the target order picker and the target case leads to a more fluent collaboration and more positive perception of the robot than a robot that directs its attention to only one of these objects, could explain a significant part of the differences between the conditions. Interestingly, the *Order picker* and *Order picker & Case* conditions did not differ significantly in terms of the average movement time, the perceived fluency of the HRC, and the perceived ease of use of the robot (only in the perceived usefulness of the robot). This, and taking into account the second hypothesis, indicates that communicating the target order picker was very important in enabling participants to anticipate the robot's behavior and that communicating the target case in addition to the target order picker had a relatively small benefit (Cohen's $d_z = 0.67$ (average movement time) / 0.40 (perceived fluency of HRC / 0.92 (perceived usefulness of the robot) / 0.28 (perceived ease of use of the robot)). These results extend the literature about the different steps involved in the process of establishing attention, as several researchers (Brinck, 2008; Huang & Thomaz, 2010; Moore et al., 2014) have described different steps, but not studied the effects of individual steps on the fluency of joint actions.

To conclude, the first step of establishing joint attention, which was used to indicate the target order picker, had the largest effect on the (perceived) fluency of the HRC and participants' perceptions of the robot in a collaborative case picking context. This finding was also supported by the results of the repeated measures ANOVA's, where the *Order picker* factor had a significant effect on all four measures (and higher values for η^2), but the *Case* factor had a significant effect only on the average movement time and perceived usefulness of the robot. The robot in the *Order picker & Case* condition was perceived as most useful by the participants. For the other measures, there was not one condition in which the robot was evaluated best (there were no significant differences between the *Order picker* and *Order picker & Case* conditions). The main results of the current study seem to be in line with the finding of a positive effect of joint attention on the fluency of the HRC and the perceptions of the robot found in static humanoid robots (Admoni et al., 2014; Boucher et al., 2012; Huang & Thomaz, 2010; Mutlu et al., 2013; Yonezawa et al., 2007). This suggests that the findings can be applied to mobile non-humanoid robots as well. However, these previous studies mainly concerned robots that used gaze cues to direct attention to a target object that represented the goal of the interaction. This behavior matches with the second step of a joint attention process. The current study showed that in a collaborative case picking scenario, the first step has a larger effect on the fluency of the HRC and the perceptions of the robot than the second step. It would be interesting to research whether the different steps involved in establishing joint attention have the same effects in different contexts.

Implications for HRI Design

The results of the current study could aid in the design of a non-humanoid mobile robot that cooperates with human partners in an industrial setting like a warehousing environment. Current designs of warehouse robots (see Figure 10 for the Locus robot[4] and Figure 11 for the Casey robot[5]), use an explicit display like a tablet for the interaction with order pickers. A code on the screen indicates the location of the target case, so the order pickers can only infer the goal of the robot when they are close enough to read the code on the screen. These robots have no way of indicating from whom they expect to receive a case.

Based on the results of the current study, the (perceived) fluency of the HRI and the perceptions of the robot could be improved by letting the robot communicate its goals by establishing joint attention with a human order picker. Although non-humanoid robots do not allow the use of gaze cues, the current study showed that alternative methods (like multi color LED lights) can be used to indicate what a robot is attending to. The first step of establishing joint attention (in which the robot makes

[4] https://locusrobotics.com/industry_solutions/industrial/

[5] https://www.gideon.ai/solutions/casey/

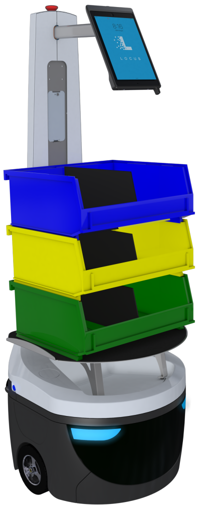

Figure 10. Locus robot developed by Locus Robotics for item picking

Figure 11. Casey robot developed by Gideon for case picking

clear that it intends to interact with a specific agent) turned out to be most effective. As this is not part of the interactions in the current designs, a large improvement could be gained. For example, the time needed to pick one case could be decreased by $(3434.42 - 1658.58)/3434.42 \times 100\% = 51.71\%$. If the robot would also direct attention to the target case, the time needed to pick one case could be decreased by $(3434.42 - 1187.50)/3434.42 \times 100\% = 65.42\%$. In addition, the order pickers' perceptions of the fluency of the HRC and the usefulness and ease of use of the robot would be more positive. However, designing a robot that only indicates the case it aims to collect by directing the attention of the order picker towards it, would be less effective. Only the movement time would decrease, but not as much as when the target order picker would be indicated. Moreover, when LED lights would be used to direct attention to a case, each unique location would require a separate LED light, which could be very costly for a large warehouse.

The current study can be viewed as the starting point for designing a non-humanoid robot that establishes joint attention. In the experiment, the positioning and colors of the LED lights were used to indicate what objects the robot was attending to. However, depending on the specific context, there might be alternative, more appropriate, methods to indicate what a robot is attending to. For example, using a set of unique colors to refer to a specific person might be impractical in an environment with more than fifteen people or people who are color-blind (as the perceptual differences between colors become smaller). An alternative method could be to use a spotlight (placed on top of the robot) that points to objects of interest. However, this method did not seem appropriate for a warehousing environment where the order pickers are standing in-between the racks and the robots. The light may be occluded or shine in a person's eyes. From the perspective of user experience design, it would be interesting to try out different methods and compare their effectiveness in establishing joint attention.

An important finding of the current study is that signals different from gaze cues (which are often used by humanoid robots) can be used by a robot to indicate its intentions. This means that robots do not necessarily have to look human in order to be able to make their intentions clear by directing an observer's attention to a target object. This is in line with the findings of Szafir et al. (2015), who used a LED ring attached to a small flying robot to enable observers to predict the target of the robot's movements. In line with the GOADI theory, the conclusion is that visual cues should convey high level aspects of a robot's intentions rather than low level aspects in order to enable observers to anticipate the robot's behavior.

Limitations and Future Research

The current study has several limitations. The first limitation is related to the external validity, which indicates the extent to which the findings of the study can be applied to other settings. The desired participants were actual order pickers, because they have experience with the case picking task. They are representative of the target user group and it might be easier for them to judge whether the robot (and its light cues) would be considered useful in a collaborative case picking task. However, due to issues related to the feasibility of the study, the participants in the experiment were not order pickers, but employees at Vanderlande. The results of the demographics questionnaire indicated that only one of the participants had experience with order picking tasks from the past. However, too little data was collected in order to be able to say something meaningful about the effect of experience on the perceptions of the robot. To keep the external validity as high as possible, the scenario was adapted accordingly. As mentioned before, the labels for the locations in a warehouse are quite complicated, but experienced order pickers are familiar with the underlying system, so the codes are clear to them and they know the general layout of the warehouse. To simulate this, the labels on the racks in the experiment were simplified and the underlying system was explained to the participants beforehand. So, the level of difficulty that order pickers and participants experienced was similar. Because order pickers are used to lifting heavy objects, they might be faster at picking up and transporting cases than the average person. Therefore, further research is needed to study whether the effect sizes found in the current study are similar for experienced order pickers.

A second limitation is the unexpected issues with controlling the LED lights during the experiment. Although the researcher had tested all twenty LED lights before the start of the experiments and during a pilot study, some of the lights on the racks did not respond to the remote during one of the experiments. Therefore, the lights were detached from the racks and one functioning LED light was moved by the experimenter in-between trials to indicate to which case the robot was attending. This light was placed in the bottom-left corner of the compartment containing the target case. Participants were told about the adaptation and they were instructed to imagine that each location had an LED light attached next to the label. Although the researcher tried to fix the issues with the dysfunctional LED lights (for example, replacing the batteries or ordering new lights), there was no guarantee that the lights would keep functioning throughout the entire experiment. Therefore, only one LED light was used in the remaining experiments to indicate what case the robot was attending to.

To conduct the experiments, the Wizard of Oz method was used. The experimenter controlled the LED lights and the movements of the robot, while in a warehousing environment, the robot is supposed to move autonomously and control its own light cues. The experimenter observed when the joint action was successful (the participant had placed the target case on the pallet carried by the robot). In practice, the robot should be able to determine this itself. In current designs of warehouse robots, human order pickers press a button on the screen presented by the robot to confirm that they picked an item or case. This does not contribute to a natural interaction. An alternative method could be (most suitable for case picking) that the robot has a build-in scale that measures the weight of the load on the pallet, so that when the scale sensor measures a substantial increase in weight, the robot assumes that the joint action was successful. To improve this system, knowledge about the weights of different types of cases could be used (for example, when multiple cases of the same type need to be picked). Additional research is needed to develop a robot that can sense when a joint action has been completed.

The final limitation is related to the steps involved in the process of establishing joint attention. The first step is used to communicate the intend to have an interaction. In the experiment, this was done by changing the color of the LED lights on the robot and assigning a unique color the each order picker. During the second step, the initiating agent directs its attention to the object of interest with the aim to draw the attention of the responding agent to this object as well. The robot used the LED lights on the racks to draw the attention of the participants to the target case. In the third and final step, the initiating agent checks whether the responding agent is indeed attending to the referred object. However, this step was not explicitly implemented during the experiment. The robot (or in fact, the experimenter) did not check whether the participants were indeed attending to the target case. Efforts have been made to develop robots that can predict referential and mutual gaze in real time (Saran, Majumdar, Shor, Thomaz, & Niekum, 2018). It would be interesting to implement such a method in future studies to test if it would make the interaction more natural. The goal of the third step is to decide whether the previous step should be repeated. To make sure that participants would notice to what object the robot was attending, the light on the racks was turned on during the whole trial (so as long as was needed for the participant to pick the case). In HHI, the previous steps are performed consecutively, while in the experiment, the first two steps were performed by the robot at the same time. Gaze cues can only direct attention to one object at a time, but light cues can be used simultaneously. One the one hand, this could improve the interaction, as more information is available at the same time, but on the other hand, it is difficult for humans to attend to two objects at different locations in physical space. More research is needed to study whether consecutive or parallel execution of the different steps leads to the most fluent interaction.

Conclusion

In order to address the research question, *How does a mobile non-humanoid robot that establishes joint attention with a human partner affect the fluency of the human-robot interaction in a collaborative case picking scenario?*, three hypotheses were tested. The first hypothesis stated that a robot that directs its attention to the target order picker and/or the target case leads to a more fluent collaboration and more positive perception of the robot than a robot that does not establish joint attention with its human partner. The second hypothesis stated that a robot that directs its attention to the target order picker leads to a more fluent collaboration and more positive perception of the robot than a robot that directs its attention to the target case. The third and final hypothesis stated that a robot that directs its attention to both the target order picker and the target case leads to a more fluent collaboration and more positive perception of the robot than a robot that directs its attention to only one of these objects. Overall, all three hypotheses were supported by the data. To be more specific, directing attention to the target order picker and/or the target case led to a significant decrease in the average movement time (compared to the baseline condition). Directing attention to only the target case did not make the participants perceive the HRC as significantly more fluent or the robot as more useful or easier to use, but directing attention to the target order picker (and target case) had a significant and positive effect on the subjective experiences of the participants. A robot that communicates the target order picker significantly improved the (perceived) fluency of the HRC and the perceived usefulness and ease of use of the robot, compared to a robot that communicates the target case. The reason for this is that the different steps of establishing joint attention have different functions. Combining the cues that indicate the target order picker and the target case, led to participants perceiving the robot as significantly more useful.

To conclude, the effect of a non-humanoid robot that establishes joint attention with a human partner depends on which steps of a joint attention process are performed. The first and second step of a joint attention process correspond to communicating different types of goals. In the first step, an agent communicates with whom it intends to interact. In the second step, an agent communicates the goal of the interaction by directing the attention of both agents to the object of interest. When an observing agent can infer the goals of the initiating agent, the observing agent can use this information to anticipate future behavior of the initiating agent. This aids in the planning and selection of appropriate actions of the observing agent, which complement the actions of the other agent. The result is a more fluent and successful joint action. The current study showed that a non-humanoid robot that communicated with whom it intended to interact (with LED lights) led to a significantly more fluent interaction (in terms of the average movement time of participants) and more positive perceptions (in terms of the perceived fluency of the HRC and the perceived usefulness and ease of use of the robot). So, being able to infer whether the robot intended to interact with a participant, made it easier for participants to determine what they should do (before the robot arrived at its destination). In addition, a robot that only performed the second step (communicating the goal of the interaction), only led to a more fluent collaboration. Apparently, the participants did not feel like they were able to infer all of the relevant information about the robot's intentions. Finally, a robot that combined and performed both steps at the same time had about an equally large effect on the fluency of the interaction and the perceptions of the robot as a robot that only performed the first step.

References

Admoni, H., Bank, C., Tan, J., Toneva, M., & Scassellati, B. (2011). Robot gaze does not reflexively cue human attention. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *33*.

Admoni, H., Datsikas, C., & Scassellati, B. (2014). Speech and gaze conflicts in collaborative human-robot interactions. *Cognitive Science*, *36*.

Anjum, M. L., Ahmad, O., Rosa, S., Yin, J., & Bona, B. (2014). Skeleton tracking based complex human activity recognition using kinect camera. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8755*, 23-33. doi: 10.1007/978-3-319-11973-1_3

Bakeman, R., & Adamson, L. B. (1984, August). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development*, *55*, 1278-1289.

Bao, Y., & Cuijpers, R. H. (2017, November). On the imitation of goal directed movements of a humanoid robot. *International Journal of Social Robotics*, *9*, 691-703. doi: 10.1007/s12369-017-0417-8

Bejerano, G., LeMasurier, G., & Yanco, H. A. (2018, March). Methods for providing indications of robot intent in collaborative human-robot tasks. In (p. 65-66). ACM. doi: 10.1145/3173386.3177069

Bekkering, H., Wohlschläger, A., & Gattis, M. (2000, February). Imitation of gestures in children is goal-directed. *Quarterly Journal of Experimental Psychology*, *53*, 153 - 164.

Belhassein, K., Fernández-Castro, V., Mayima, A., Clodic, A., Pacherie, E., Guidetti, M., ... Cochet, H. (2022, February). Addressing joint action challenges in hri: Insights from psychology and philosophy. *Acta Psychologica*, *222*, 103476. doi: 10.1016/j.actpsy.2021.103476

Boucher, J. D., Pattacini, U., Lelong, A., Bailly, G., Elisei, F., Fagel, S., ... Ventre-Dominey, J. (2012, May). I reach faster when i see you look: Gaze effects in human-human and human-robot face-to-face cooperation. *Frontiers in Neurorobotics*, *6*, 3-3. doi: 10.3389/FNBOT.2012.00003

Brinck, I. (2008). *The role of intersubjectivity in the development of intentional communication.* John Benjamins Publishing. doi: 10.1075/celcr.12.08bri

Cuijpers, R. H., Schie, H. T., Koppen, M., Erlhagen, W., & Bekkering, H. (2006, April). Goals and means in action observation: A computational approach. *Neural Networks*, *19*, 311-322. doi: 10.1016/j.neunet.2006.02.004

Curioni, A., Knoblich, G., & Sebanz, N. (2017). *Joint action in humans: A model for human-robot interactions.* Springer Netherlands. doi: 10.1007/978-94-007-7194-9_126-1

Davis, F. D. (1989, September). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*, 319. doi: 10.2307/249008

Droeschel, D., Stückler, J., Holz, D., & Behnke, S. (2011). Towards joint attention for a domestic service robot - person awareness and gesture recognition using time-of-flight cameras. *Proceedings - IEEE International Conference on Robotics and Automation*, 1205-1210. doi: 10.1109/ICRA.2011.5980067

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods 2007 39:2*, *39*, 175-191. doi: 10.3758/BF03193146

Haans, A. (2018, January 1). Contrast analysis: A tutorial. *Practical Assessment, Research Evaluation*, *23*(9), 1–21. doi: 10.7275/zeyh-j468

Hoffman, G., & Breazeal, C. (2007, October). Cost-based anticipatory action selection for human-robot fluency. *IEEE Transactions on Robotics*, *23*, 952-961. doi: 10.1109/TRO.2007.907483

Huang, C.-M., & Thomaz, A. L. (2010). Joint attention in human-robot interaction. In *Aaai fall symposium: Dialog with robots.*

Luo, Y., & Baillargeon, R. (2005). Can a self-propelled box have a goal?: Psychological reasoning in 5-month-old infants. *Psychological Science*, *16*(8), 601-608. doi: 10.1111/j.1467-9280.2005.01582.x

Mainprice, J., & Berenson, D. (2013). Human-robot collaborative manipulation planning using early prediction of human motion. In (p. 299-306). doi: 10.1109/IROS.2013.6696368

Moore, C., Dunham, P. J., & Dunham, P. (2014). *Joint attention : Its origins and role in development.* Psychology Press. doi: 10.4324/9781315806617

Movellan, J. R., & Watson, J. S. (2002). The development of gaze following as a bayesian systems identification problem. *Proceedings 2nd International Conference on Development and Learning. ICDL 2002*, 34-40.

Mutlu, B., Huang, C.-M., & Terrell, A. (2013). Coordination mechanisms in human-robot collaboration. In (p. 1-6).

Paliga, M., & Pollak, A. (2021, November). Development and validation of the fluency in human-robot interaction scale. a two-wave study on three perspectives of fluency. *International Journal of Human Computer Studies*, *155*. doi: 10.1016/J.IJHCS.2021.102698

Saran, A., Majumdar, S., Shor, E. S., Thomaz, A., & Niekum, S. (2018, December). Human gaze following for human-robot interaction. *IEEE International Conference on Intelligent Robots and Systems*, 8615-8621. doi: 10.1109/IROS.2018.8593580

Sauppé, A., & Mutlu, B. (2014, March). Robot deictics. In *Proceedings of the 2014 ACM/IEEE international conference on human-robot interaction.* ACM. doi: 10.1145/2559636.2559657

Sebanz, N., Bekkering, H., & Knoblich, G. (2006, February). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*, 70-76. doi: 10.1016/J.TICS.2005.12.009

Sharma, R. (2019, March). *True ventures - give robots the repetitive, tedious work.* Retrieved from `https://trueventures.com/blog/give-robots-the-repetitive-work`

Szafir, D., Mutlu, B., & Fong, T. (2015, March). Communicating directionality in flying robots. *ACM/IEEE International Conference on Human-Robot Interaction*, *2015-March*, 19-26. doi: 10.1145/2696454.2696475

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). *User acceptance of information technology: Toward a unified view* (Vol. 27).

Wohlschläger, A., Gattis, M., & Bekkering, H. (2003, March). Action generation and action perception in imitation: an instance of the ideomotor principle. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*, 501-515. doi: 10.1098/rstb.2002.1257

Yekutieli, Y., Sagiv-Zohar, R., Aharonov, R., Engel, Y., Hochner, B., & Flash, T. (2005, August). Dynamic model of the octopus arm. i. biomechanics of the octopus reaching movement. *Journal of neurophysiology*, *94*, 1443-1458. doi: 10.1152/JN.00684.2004

Yekutieli, Y., Sagiv-Zohar, R., Hochner, B., & Flash, T. (2005, August). Dynamic model of the octopus arm. ii. control of reaching movements. *Journal of neurophysiology*, *94*, 1459-1468. doi: 10.1152/JN.00685.2004

Yonezawa, T., Yamazoe, H., Utsumi, A., & Abe, S. (2007). Gaze-communicative behavior of stuffed-toy robot with joint attention and eye contact based on ambient gaze-tracking. In *International conference on multimodal interaction.*

Appendix A

Informed Consent Form

VANDERLANDE   TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Information sheet for research project "Collaborative case picking with a mobile robot"

1. Introduction

You have been invited to take part in the research project Collaborative case picking with a mobile robot, because you were selected from the Vanderlande employee pool.

Participation in this research project is voluntary: the decision to take part is up to you. Before you decide to participate we would like to ask you to read the following information, so that you know what the research project is about, what we expect from you and how we go about processing your personal data. Based on this information you can indicate by way of the consent declaration whether you consent to taking part in this research project and in the processing of your personal data.

You may of course always contact the researcher via Nynke.Boonstra@vanderlande.com, if you have any questions, or you can discuss this information with people you know.

2. Purpose of the research

This research project will be managed by dr. ir. Raymond Cuijpers (R.H.Cuijpers@tue.nl) and Linda van der Meijden (Linda.van.der.Meijden@vanderlande.com). The research project is a collaboration between TU/e and Vanderlande.
The primary purpose of this research project is to evaluate the interaction between a mobile robot and a human order picker in a collaborative case picking scenario. This will provide insights on how to design a cooperative mobile robot that can be used by Vanderlande to make case picking more efficient. The research data will be used for a master thesis.
Additionally, the TU/e might process the anonymized data for related research projects in the future.

3. Controller in the sense of the GDPR

The research team is responsible for processing your personal data within the scope of the research on behalf of joint controllers TU/e and Vanderlande.

4. What project will you be taking part in within the research project involved?

You will be taking part in a research project in which we will gather information by:

- Observing your collaboration and interaction with the robot (and making video recordings of this).
- Asking you to fill in several questionnaires about your experiences with the robot and demographics.

For your participation in this research project you will not be compensated.

5. What personal data from you do we gather and process?

Within the framework of the research project we process the folllowing personal data:

Category	Personal data	Purpose
Demographic data	*Name, age and gender*	*Your name is processed so that we can show whether you gave consent.* *Age and gender are processed to analyze the background of different groups of participants.* *TU/e might use the anonymized demographical data (so not your name) for related research projects in the future.*
Video data	*Physical movements of the body during interaction with the robot*	*Analyze data concerning your movement behavior during the interaction with the robot.* *TU/e might use the anonymized data obtained from the video recordings for related research projects in the future.*

6. Withdrawing your consent and contact data

Participation in this research project is entirely voluntary. You do not have to answer questions you do not wish to answer. You may end your participation in the research project at any moment, or withdraw your consent to using your data for the research, without specifying any reason. Ending your participation will have no disadvantageous consequences for you.

If you decide to end your participation during the research, the data which you already provided up to the moment of withdrawal of your consent will be used in the research.
Do you wish to end the research, or do you have any questions and/or complaints? Then please contact the researcher via Nynke.Boonstra@vanderlande.com.

If you have specific questions about the handling of personal data you can direct these to the data protection officer of Vanderlande by sending a mail to geert.van.rooy@vanderlande.com. Furthermore, you have the right to file complaints with the Dutch data protection authority: the Autoriteit Persoonsgegevens.

Finally, you have the right to request access, rectification, erasure or adaptation of your data. Submit your request via Nynke.Boonstra@vanderlande.com.

7. Legal ground for processing your personal data

To be permitted to process your personal data, the processing must be based on one of the legal bases from the GDPR. For this research project, in which we record video footage to register collaborative case picking with a mobile robot, explicit consent is the legal ground.
Regarding your participation in this research project, Vanderlande has a legitimate interest to invite you to cooperate with this project: Vanderlande needs to test and optimize the design of a collaborative mobile robot.

8. Who has access to your personal data?

Access to personal data within TU/e and Vanderlande
The research team has access to your personal data, but only as far as is necessary to fulfil their respective tasks. The research team consists of a master student (Nynke Boonstra), two supervisors at TU/e (dr. ir. R. H. Cuijpers and M. M. E. Neggers) and one supervisor at Vanderlande (L. van der Meijden). Beside this research team, only authorized persons in the relevant sections of TU/e and Vanderlande like engineers at Vanderlande that supervise the use of the robot and the camera setup will have access to your personal data, but only as far as is necessary to fulfil their respective tasks. Although the video footage will not be directly linked to your name, you might be recognisable.

The data collected in this study might also be of relevance for future research projects within the Human Technology Interaction group as well as for other researchers. The aim of those studies might be unrelated to the goals of this study. The collected data will therefore also be made available to authorized researchers from other institutions, but not the general public, in an online data repository of TU/e with restricted access. The coded data collected in this study and that will be stored in an online data repository (this includes only anonymized data, so no actual video recordings) will (to the best of our knowledge and ability) not contain information that can identify you.

9. How are your personal data protected?

TU/e and Vanderlande have implemented appropriate technical and organizational measures for protection of personal data against unintended or unlawful destruction, unintended damage, loss, alteration and unauthorized publication or access, and against all other forms of unlawful processing (including, but not limited to unnecessary gathering of data) or further processing. These appropriate technical and organizational measures include storing the data on a protected server, limating the access to the data through authorization and authentication, and removing the video recordings after (anonymous) data regarding the physical movement behavior has been deduced. Data minimalization and retention time has been minimized.

10. How long will your personal data be retained?

Your personal data will be retained in accordance with the GDPR. The data are retained no longer than is necessary to achieve the goals for which the data were gathered and are deleted as soon as you withdraw your consent and there is no other ground to process your data lawfully. The video recordings are temporarily stored on an encrypted laptop of Vanderlande and are deleted at the end of the project (about 10 weeks after the start of the experiment). The signed informed consent forms will be stored until the project has been completed, reported and assessed and an additional four weeks to mitigate legal disputes. Vanderlande will store digital copies, while the physical copies will be stored by TU/e. Your anonymized data are deleted after no more than 10 years by TU/e.

11. Confidentiality of data

We will do everything we can to protect your privacy as best as possible. The research results that are published will in no way contain confidential information or personal data from or about you through which anyone can recognize you, unless you have by way of our consent form explicitly consented to mentioning your name, for example in a quote. The research data will if necessary (for example for a check on scientific integrity) and only in anonymized form be made available to people outside the research group. All members of the research team have signed a non-disclosure agreement.

If you consent, your anonymized data will be made available for future research at the TU/e via Mircrosoft OneDrive.

Finally, this research has been assessed and approved by the ethical committee of Eindhoven University of Technology.

VANDERLANDE

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Consent form for participation

By signing this consent form I acknowledge the following:

1. I am sufficiently informed about the research project through a separate information sheet. I have read the information sheet and have had the opportunity to ask questions. These questions have been answered satisfactorily.

2. I take part in this research project voluntarily. There is no explicit or implicit pressure for me to take part in this research project. I am clear that I can end participation in this research project at any moment, without giving any reason. I do not have to answer a question if I do not wish to do so.

Furthermore, I consent to the following parts of the research project

3. I consent to processing my personal data gathered during the research in the way described in the information sheet.

   YES ☐ NO ☐

4. I consent to making video recordings of the experiment.

   YES ☐ NO ☐

5. I consent to retaining anonymized research data gathered from me in an online repository of TU/e and using this for future research in the field of collaborative case picking and human-robot interaction in which recognized ethical standards for scientific research are respected, and for education purposes.

   YES ☐ NO ☐

Name of Participant:

Signature:

Date:

Name of Researcher:

Signature:

Date:

Appendix B

Participant Instructions

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

VANDERLANDE

Dear participant,

Thank you for participating in this experiment. Please read the following instructions carefully.

Imagine that you are an order picker working in a warehousing environment. Because Christmas is coming up, the number of cases that need to be picked is very high. You work fast, while maintaining a high level of accuracy. You collaborate with a mobile stacker robot to collect cases for several orders. The robot displays a code that matches with the location of a case on the racks. Your task is to pick up the case at the specified location and put it on the pallet carried by the robot (as fast as possible). The exact position of the case is not important, but the cases should not fall off the pallet.

One picking task consists of the following steps:

1. At the start, you stand behind a table and face a fence.
2. When the experimenter says "Start", you start walking towards the racks.
3. You pick up the case at the location that is indicated by the code provided by the approaching robot.
4. You put the case on the pallet carried by the robot.
5. The robot continues its path, and you walk back to the table and get ready for the next picking task.

Because multiple robots and order pickers are working in the same environment, you sometimes encounter robots that pass you because they have been given the instruction to collaborate with another human order picker. You do not collect cases for these robots.

The robot provides several signals for safety reasons, but it also tries to communicate its goals. The color of the LED lights attached to the robot indicates the order picker it needs to interact with. Every order picker is assigned a unique color. You are assigned the color orange. The robot indicates the case it needs to collect by turning on a LED light below this case (white means neutral, a color also indicates a certain order picker).

The experiment consists of 4 sessions of picking tasks, with a short survey after each session.

If you have any questions, please inform the researcher.

Beste participant,

Bedankt voor je deelname aan dit experiment. Lees de volgende instructies aandachtig.

Beeld je in dat je een orderverzamelaar bent die in een warenhuis werkt. Omdat kerst eraan komt, moeten er erg veel pakketten worden verzameld. Daarom werk je hard, maar ook nauwkeurig. Je werkt samen met een stapelaar robot om pakketten voor orders te verzamelen. De robot laat een code zien die overeenkomt met de locatie van een pakket op de rekken. Het is jouw taak om het pakket op de aangegeven locatie op te pakken en het op de pallet te plaatsen die wordt gedragen door de robot. Het is niet belangrijk hoe je het pakket precies neerzet, maar het mag niet van de pallet vallen.

Eén verzamelopdracht bestaat uit de volgende stappen:

1. Aan het begin sta je achter een tafel met je gezicht naar een hek gericht.
2. Als de onderzoeker “Start” zegt, begin je naar de rekken te lopen.
3. Je pakt het pakket op dat op de locatie staat die wordt aangegeven door de code op de zijkant van de naderende robot.
4. Je plaatst het pakket op de pallet die door de robot wordt gedragen.
5. De robot gaat verder met zijn route en jij loopt terug naar de tafel en maakt je klaar voor de volgende verzamelopdracht.

Omdat meerdere robots en orderverzamelaars in dezelfde omgeving werken, zul je soms robots tegenkomen die je passeren omdat ze de instructie hebben gekregen om met een andere orderverzamelaar samen te werken. Je hoeft voor deze robots geen pakketten te verzamelen.

De robot geeft een aantal signalen vanwege veiligheidsredenen, maar hij probeert ook om zijn doelen te communiceren. De kleur van de LED lampjes op de robot geeft aan met welke orderverzamelaar de robot een interactie wil aangaan. Elke orderverzamelaar heeft een unieke kleur toegewezen gekregen. Aan jou is de kleur oranje toegewezen. De robot geeft aan welk pakket hij wil verzamelen door een LED lampje net onder het pakket aan te zetten (wit licht is neutraal, een kleur geeft ook de orderverzamelaar aan).

Het experiment bestaat uit 4 sessies orders verzamelen en een korte vragenlijst na elke sessie.

Als je nog vragen hebt, stel ze dan aan de onderzoeker.

Appendix C

Subjective Experiences Questionnaire

Participant ID: Session number:

Based on your experience with the robot in the previous eight picking tasks, please indicate whether you agree or disagree with the following statements.

Perceived fluency of the human-robot collaboration

	I strongly disagree	I disagree	I somewhat disagree	I neither agree nor disagree	I somewhat agree	I agree	I strongly agree
I trusted the robot to do the right thing at the right time.	○	○	○	○	○	○	○
I felt like the robot was committed to the success of the team.	○	○	○	○	○	○	○
The robot performed well as part of the team.	○	○	○	○	○	○	○
The robot did its part successfully.	○	○	○	○	○	○	○
The human-robot team did well on the task.	○	○	○	○	○	○	○
The human-robot team felt well-tuned.	○	○	○	○	○	○	○

Please see the other side for additional questions

Perceived usefulness & ease of use of the robot

	I strongly disagree	I disagree	I somewhat disagree	I neither agree nor disagree	I somewhat agree	I agree	I strongly agree
Using the robot in a case picking task enables me to accomplish picking tasks more quickly.	○	○	○	○	○	○	○
Using the robot improves my case picking performance.	○	○	○	○	○	○	○
Using the robot in a case picking task increases my productivity.	○	○	○	○	○	○	○
Using the robot enhances my effectiveness in a case picking task.	○	○	○	○	○	○	○
Using the robot makes it easier to execute case picking tasks.	○	○	○	○	○	○	○
I find the robot useful in a case picking task.	○	○	○	○	○	○	○

	I strongly disagree	I disagree	I somewhat disagree	I neither agree nor disagree	I somewhat agree	I agree	I strongly agree
Learning to collaborate with the robot is easy for me.	○	○	○	○	○	○	○
My interaction with the robot is clear.	○	○	○	○	○	○	○
It is easy for me to become skillful at collaborating with the robot.	○	○	○	○	○	○	○
I find the robot easy to collaborate with.	○	○	○	○	○	○	○
My interaction with the robot is understandable.	○	○	○	○	○	○	○

Appendix D

Demographics Questionnaire

Demographics

Please answer the following questions about your demographics.

* Required

* This form will record your name, please fill your name.

1. What is your participant ID? (please ask the researcher) *

2. What is your age? *

The value must be a number

3. What is your gender? *

- Female
- Male
- Prefer not to say / other

4. Please describe (in short) your previous experiences with mobile robots (before your participation in this experiment). If you do not have any experience with such robots, please indicate this. *

5. Please describe (in short) your previous experiences with order picking (before your participation in this experiment). If you do not have any experience with this kind of work, please indicate this. *

Appendix E

Line Graphs

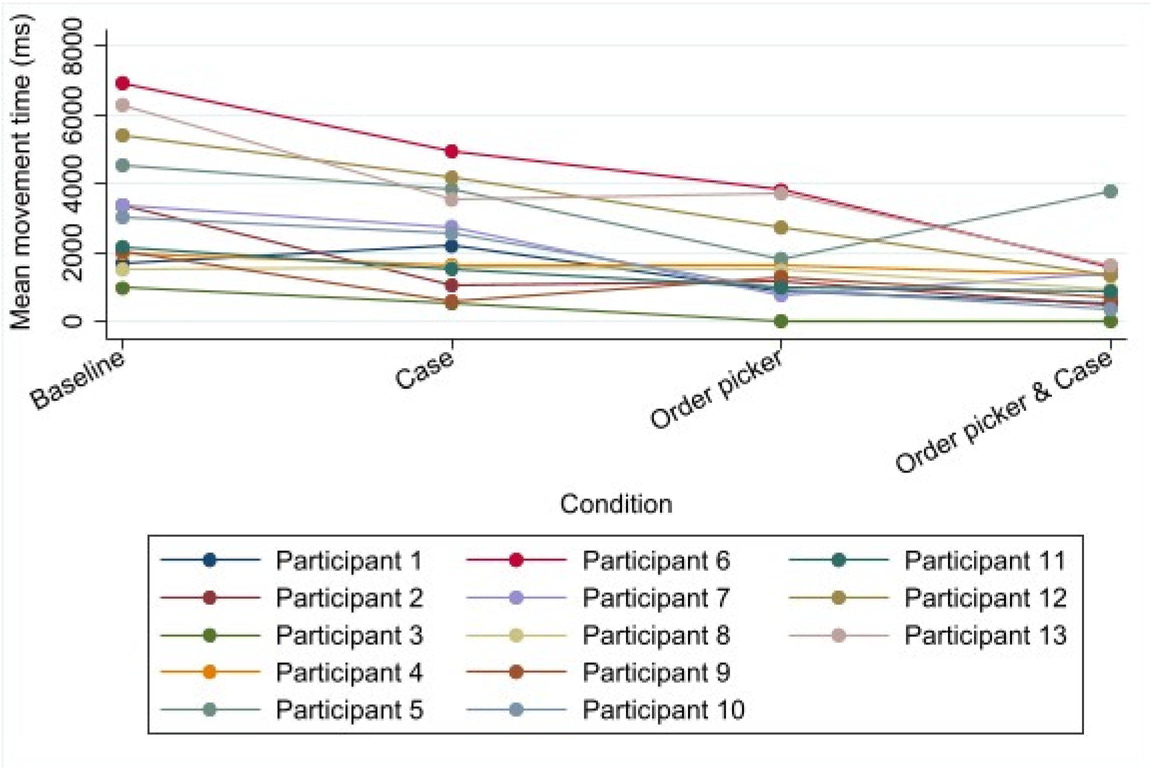

Figure E1. Individual data points representing the average movement time (ms) per condition per participant. Data points belonging to the same participant are connected with a line and assigned the same color.

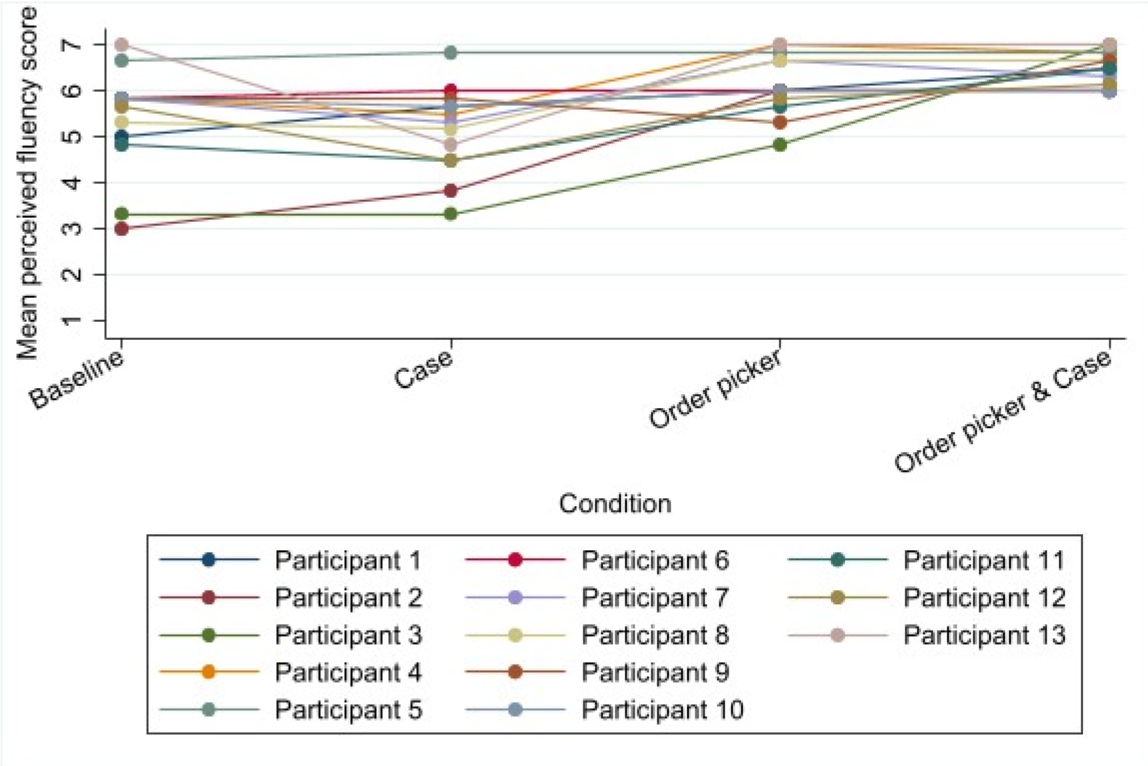

Figure E2. Individual data points representing the perceived fluency of the HRC per condition per participant. Data points belonging to the same participant are connected with a line and assigned the same color.

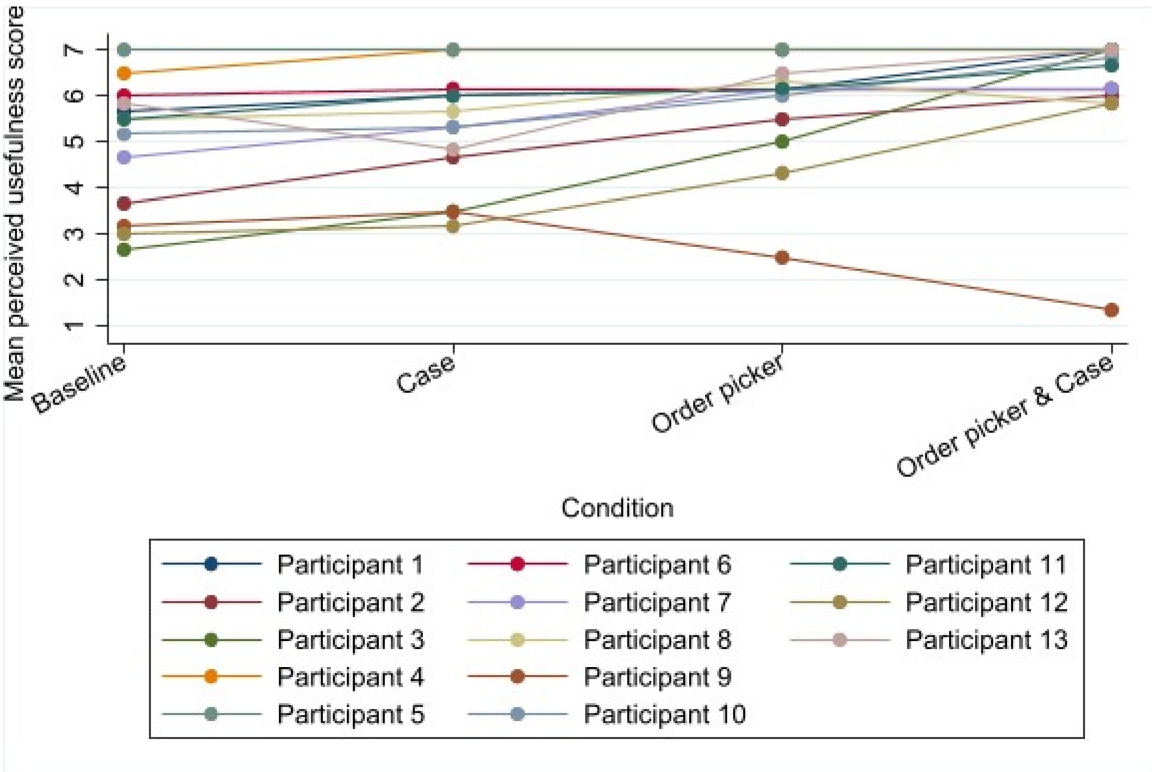

Figure E3. Individual data points representing the perceived usefulness of the robot per condition per participant. Data points belonging to the same participant are connected with a line and assigned the same color. The data of participant 9 deviates from the overall pattern.

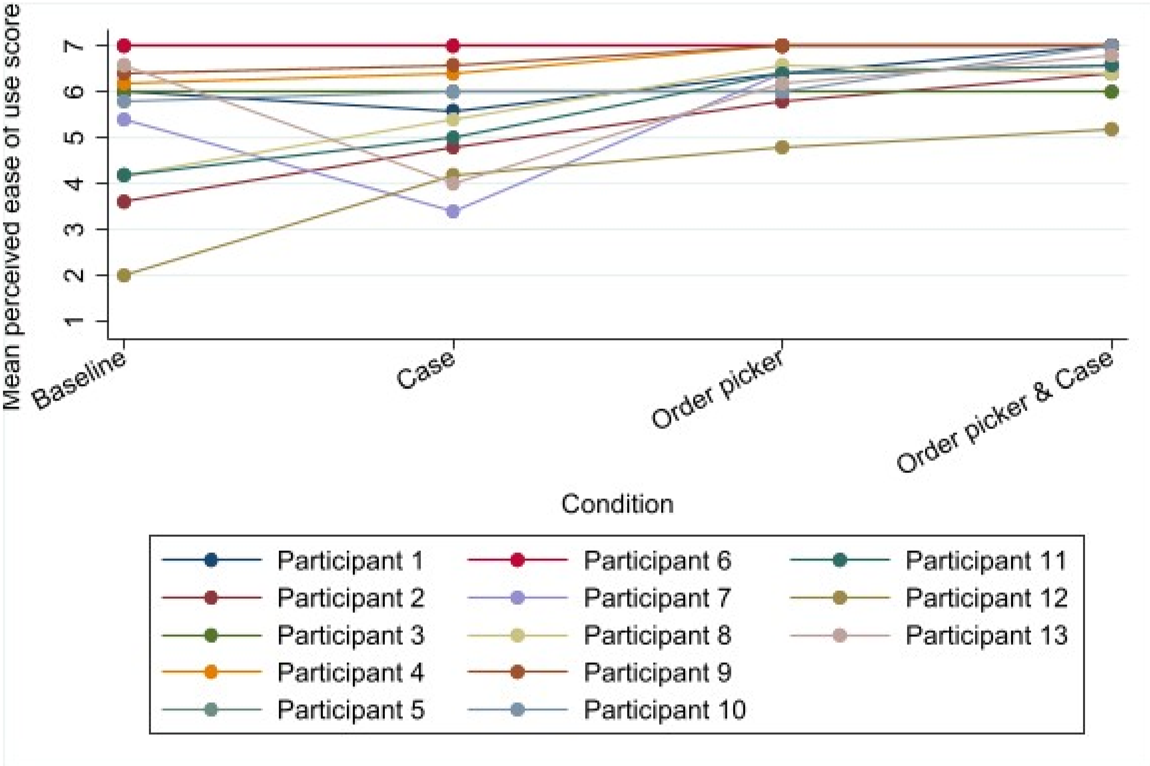

Figure E4. Individual data points representing the perceived ease of use of the robot per condition per participant. Data points belonging to the same participant are connected with a line and assigned the same color.

Appendix F

Exploratory Analysis

Table F1

Correlations between variables for each condition

Variable	Condition	1	2	3	4
1. Average movement time (ms)	Baseline	-			
	Case	-			
	Order picker	-			
	Order picker & Case	-			
2. Perceived fluency of the HRC	Baseline	.55	-		
	Case	.60	-		
	Order picker	.41	-		
	Order picker & Case	.24	-		
3. Perceived usefulness of the robot	Baseline	.17	.67	-	
	Case	.12	.77*	-	
	Order picker	.12	.73*	-	
	Order picker & Case	.11	.59	-	
4. Perceived ease of use of the robot	Baseline	.13	.39	.62	-
	Case	.07	.46	.53	-
	Order picker	.06	.46	.89***	-
	Order picker & Case	.24	.09	.53	-

Note. * $p < .05$, ** $p < .01$, *** $p < .001$ after Bonferroni correction.

Table F2

Results of two-way repeated measures ANOVA's

Variable	Factor	*df*	*MS*	*F*	*p*	η^2
Average movement time (ms)	Case	1	5,682,192.20	6.50	.027	37.13%
	Order picker	1	29,157,978.00	33.84	< .001	75.47%
	Interaction	1	565,285.02	2.26	.161	17.02%
	Participant	11	6,054,990.70	7.03	.002	87.54%
Perceived fluency of the HRC	Case	1	.002	0.010	.932	0.07%
	Order picker	1	15.19	22.25	< .001	66.92%
	Interaction	1	0.84	3.49	.089	24.07%
	Participant	11	1.67	2.45	.077	70.98%
Perceived usefulness of the robot	Case	1	1.95	8.82	.013	44.50%
	Order picker	1	13.02	17.80	.001	61.80%
	Interaction	1	0.15	0.94	.352	7.90%
	Participant	11	3.01	4.11	.014	80.45%
Perceived ease of use of the robot	Case	1	0.37	0.84	.380	7.08%
	Order picker	1	13.87	20.72	< .001	65.32%
	Interaction	1	0.14	0.28	.607	2.48%
	Participant	11	2.76	4.12	.014	80.46%

Table F3

Results of Tukey's HSD post-hoc test for average movement time

Conditions	Contrast	*SE*	*Tukey t*	*p*	Cohen's d_z
Baseline vs Case	905.17	284.75	4.43	.005	1.28
Baseline vs Order picker	1775.83	304.20	8.69	$< .001$	2.51
Baseline vs Order picker & Case	2246.92	483.40	11.00	$< .001$	3.18
Case vs Order picker	870.67	236.02	4.26	.006	1.23
Case vs Order picker & Case	1341.75	304.68	6.57	$< .001$	1.90
Order picker vs Order picker & Case	471.08	326.27	2.31	.156	0.67

Note. Cohen's d_z refers to the standardized mean difference effect size for within-subjects designs.

Table F4

Results of Tukey's HSD post-hoc test for perceived fluency of the HRC

Conditions	Contrast	*SE*	*Tukey t*	*p*	Cohen's d_z
Baseline vs Case	-0.25	0.23	-1.25	.610	0.36
Baseline vs Order picker	0.86	0.25	4.31	.006	3.46
Baseline vs Order picker & Case	1.14	0.34	5.70	.001	1.65
Case vs Order picker	1.11	0.22	5.56	.001	1.60
Case vs Order picker & Case	1.39	0.31	6.95	$< .001$	2.01
Order picker vs Order picker & Case	0.28	0.19	1.39	.530	0.40

Note. Cohen's d_z refers to the standardized mean difference effect size for within-subjects designs.

Table F5

Results of Tukey's HSD post-hoc test for perceived usefulness of the robot

Conditions	Contrast	*SE*	*Tukey t*	*p*	Cohen's d_z
Baseline vs Case	0.29	0.15	1.80	.322	0.52
Baseline vs Order picker	0.93	0.20	5.75	.001	1.66
Baseline vs Order picker & Case	1.44	0.36	8.93	$< .001$	2.58
Case vs Order picker	0.64	0.17	3.95	.010	1.14
Case vs Order picker & Case	1.15	0.33	7.13	$< .001$	2.06
Order picker vs Order picker & Case	0.51	0.20	3.18	.038	0.92

Note. Cohen's d_z refers to the standardized mean difference effect size for within-subjects designs.

Table F6

Results of Tukey's HSD post-hoc test for perceived ease of use of the robot

Conditions	Contrast	*SE*	*Tukey t*	*p*	Cohen's d_z
Baseline vs Case	0.07	0.38	0.23	.995	0.07
Baseline vs Order picker	0.97	0.33	3.34	.029	0.96
Baseline vs Order picker & Case	1.25	0.33	4.32	.006	1.25
Case vs Order picker	0.90	0.27	3.11	.042	0.90
Case vs Order picker & Case	1.18	0.30	4.09	.008	1.18
Order picker vs Order picker & Case	0.28	0.10	0.98	.764	0.28

Note. Cohen's d_z refers to the standardized mean difference effect size for within-subjects designs.