



# Perception–Intention–Action Cycle in Human–Robot Collaborative Tasks: The Collaborative Lightweight Object Transportation Use-Case

J. E. Domínguez-Vidal<sup>1,2</sup> · Nicolás Rodríguez<sup>1,2</sup> · Alberto Sanfeliu<sup>1,2</sup>

Accepted: 12 January 2024  
© The Author(s) 2024

## Abstract

This study proposes to improve the reliability, robustness and human-like nature of Human–Robot Collaboration (HRC). For that, the classical Perception–Action cycle is extended to a Perception–Intention–Action (PIA) cycle, which includes an Intention stage at the same level as the Perception one, being in charge of obtaining both the implicit and the explicit intention of the human, opposing to classical approaches based on inferring everything from perception. This complete cycle is exposed theoretically including its use of the concept of Situation Awareness, which is shown as a key element for the correct understanding of the current situation and future action prediction. This enables the assignment of roles to the agents involved in a collaborative task and the building of collaborative plans. To visualize the cycle, a collaborative transportation task is used as a use-case. A force-based model is designed to combine the robot’s perception of its environment with the force exerted by the human and other factors in an illustrative way. Finally, a total of 58 volunteers participate in two rounds of experiments. In these, it is shown that the human agrees to explicitly state their intention without undue extra effort and that the human understands that this helps to minimize robot errors or misunderstandings. It is also shown that a system that correctly combines inference with explicit elicitation of the human’s intention is the best rated by the human on multiple parameters related to effective Human–Robot Interaction (HRI), such as perceived safety or trust in the robot.

**Keywords** Physical human–robot interaction · Human–robot teaming · Human-in-the-loop · User study

## 1 Introduction

In this article the classic Perception–Action (PA) cycle [1, 2] is brought up-to-day in a new framework that emphasizes the importance for each agent to know and understand the intention of its partner in collaborative tasks, that is, for the robot to know the intention of the human but also for the human to know the intention of the robot.

For this purpose, we divide all the tasks typically assigned to the classic Perception block into two blocks at the same

level. One of them is in charge of perceiving all the information necessary to understand the environment and the other one is in charge of receiving (1) directly the intention of the other agent in case they explicitly indicate it and (2) all the information necessary to infer the implicit intention from their actions. This is how the Perception–Intention–Action (PIA) cycle arises.

Its usefulness in an application that requires Human–Robot Collaboration (HRC) lies in a greater modularity that helps to know which stage is being worked on and how each one of them interrelates with the others. Likewise, and due to its generality, it can serve as a compass for a wide variety of tasks as will be shown throughout the article. In addition, it also lays on the table the possibility usually neglected in the literature that the human (or robot) can explicitly indicate their intention without the need to use complex inference systems.

To understand this small shift in thinking, it is worth looking at how we, humans, approach some of our daily tasks as these have always served as inspiration for robotics. For example, when we walk down the street, we use our sight and

---

✉ J. E. Domínguez-Vidal  
jdominguez@iri.upc.edu  
Nicolás Rodríguez  
nrodriguez@iri.upc.edu  
Alberto Sanfeliu  
alberto.sanfeliu@upc.edu

<sup>1</sup> Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Llorens i Artigas 4-6, 08028 Barcelona, Spain

<sup>2</sup> Universitat Politècnica de Catalunya - BarcelonaTech (UPC), Jordi Girona, 31, 08034 Barcelona, Spain

hearing to perceive possible obstacles or traffic signs and with these data we infer whether we should stop at a crosswalk or whether we should brake or turn slightly to one side to avoid colliding with another pedestrian. However, when it comes to performing a collaborative task with other human, we use language to communicate explicitly with our peers and thus coordinate with our workmates, obtain information from our friends or agree with our partner.

Using the above example, a robot navigating autonomously along a street will use its sensors to detect the obstacles present, infer their speed and acceleration if they are mobile and, with all this information, it will trace the route or sequence of actions it must perform to reach its destination. However, when it comes to making a robot collaborate with a human, we have not followed the above approach of imitating the human behaviour. What we have done is to use more and better sensors and more complex and powerful inference systems in order to make better and better inferences of their intent [3–5]. Interestingly, this is also a human behaviour. Whether out of fear, embarrassment or disregard of the consequences of making a mistake, we often try to infer the intentions of our fellow humans from their actions, resulting in multiple errors and misunderstandings.

This is due to the fact that the correct understanding of the intention of the other agent is essential for the correct development of this type of tasks and that this intention is not always inferable or, alternatively, that the associated uncertainty is too high. Just as we eventually come to understand that the best way to know another person's intentions, preferences and desires is to explicitly talk to them, we propose to use the same approach to make Human–Robot Collaboration (HRC) more reliable, robust and, ultimately, more human-like.

Specifically, we propose the PIA cycle by which we give the human the possibility to express their intention explicitly and we separate everything referring to the perception of the environment from the information necessary to understand the human's intention including an Intention stage at the same level of the Perception one and not as a sub-block of this. This new block can analyze jointly the implicit intention inferred through the perception of the human partner and the explicit one indicated by this human. Once the environment has been perceived and the partner's intention has been analyzed, both types of information should be properly combined and understood. For this purpose, we resort to the concept of Situation Awareness [6] (SA).

Thus, our first contribution is the statement of our theoretical framework explaining the previous division and using the concept of Situation Awareness to allow the robot to understand the current situation in which it finds itself, thus being able to adapt to it by choosing the most appropriate strategy in each case. This cycle is what allows the robot to understand the action to be taken, opening the door for it to be



**Fig. 1** Example of Human–robot pair collaboratively transporting an object. Both agents must navigate through a complex environment with multiple walls. Human has extra information since he is the only one to understand the forbidden pass sign. The transported object is a steel bar

adaptive, anticipatory or even proactive depending on the situation. This cycle also serves to explain concepts such as the cooperative roles or negotiation.

To illustrate our theoretical framework in an easy to understand yet effective manner, we chose a human–robot collaborative transportation task as a first use case. The choice of this case is due to the fact that it is a task that occurs in close proximity between human and robot, where some response speed is required limiting the complexity of the inference from the environment and the decision making. It also allows us to introduce multiple scenarios in which, for example, one of the agents has partial information or in which a specific collaboration is necessary to overcome an obstacle (see Fig. 1). Additionally, it is a case widely studied in the literature [7–13], allowing us to compare state-of-the-art solutions (typically based on inferring the human's intention through the force exerted on the object) with our approach of combining both types of intention.

In order to perform real experiments, we have designed a simple but functional force-based model based on the known Social Force Model (SFM) [14]. This model is used to represent the environment perceived by the robot as a set of attractive and repulsive virtual forces, just as [14] does representing the goal of the navigation as an attractor and each detected obstacle as a repulsor. The choice of this approach is due to the fact that it greatly simplifies the integration of the physical force exerted by the human, as well as the representation of their intention as another virtual force. Taking advantage of the model, we built a shared control system that allows us to directly combine the robot's preferences (its optimization criteria) with those of the human, being the formulation of this force-based model our second contribution.

Once our shared control system is implemented, we used it to validate our theoretical framework performing two rounds of experiments. The first round is used to check if the user accepts our framework, i.e., that it does not impose an exces-

sive extra burden and that they understand its usefulness. Later, the second round is used to compare a state-of-the-art solution based on using a complex system to infer the human's intention (in this case based on a deep learning architecture), with our framework and check if it can improve any aspect associated to an effective Human–Robot interaction (HRI).

In summary, the key findings of our work are the following. First, multiple tasks can fit into our framework, demonstrating its generality. Second, we have experimentally verified that humans appreciate having some way to explicitly indicate their intention and that this does not involve any extra physical or mental effort. Third, we have also found that humans prefer a system in which their intention is elicited both implicitly and explicitly. Thus, a system in which the robot shows enough intelligence to roughly understand the human's intention, but also allows the human to explicitly indicate its desires in order to avoid possible errors or misunderstandings is the preferred one.

Two points are worth mentioning. First, that the idea of jointly considering the human's implicit and explicit intention to improve performance in human–robot collaborative tasks in terms of parameters such as human comfort or trust in the robot was briefly presented in our previous work [15]. Albeit without the degree of depth with which it is presented here, nor with an in-depth analysis of the experimental results. In addition, the formulation of the force model was completely missing. Second, that our force model is not intended to be perfect but to visualize all the theoretical concepts we present in a simple way. This implies that any of its parts can be improved by other researchers using more complex architectures or extra modules. It is not our aim to show a definitive method that definitively solves the task analyzed here but to show another way to solve it that can serve as inspiration for other tasks.

In the remainder of the article, Section 2 presents the work related to this article. Section 3 explains our complete cycle inspired by everyday examples that can be applied to robotics. Section 4 includes the formulation of the force-based model used in order to test various parts of our cycle in real experiments. It also includes some implementation details to increase the reproducibility of our results. Section 5 presents the two rounds of experiments performed with their respective results. Finally, Section 6 discusses some of the limitations of the present work and Section 7 presents the conclusions.

## 2 Related Work

Since the beginning of robotics, the Perception–Action cycle has served as an inspiration to allow, based on how the human brain works [16], the decomposition of the robot control into

its functional modules [1, 2, 17] and the subsequent development of more advanced architectures and more complex robots [18]. This has allowed us to advance from the first machines with some autonomy [19, 20] to today's humanoid robots [21–23], improving in the process the robotic capabilities to perform specific tasks. However, despite its proven validity in improving the capabilities of an autonomous robot, the authors also conclude that this cycle is not sufficient when it comes to performing collaborative tasks with another human [24]. This is why we begin to take into account the intention of the other agent.

Examples of attempting to infer human intent using different models are common in the literature [7, 25–30]. They use the human's previous hand or whole body motion to predict the following trajectory and with this infer their intention. The analysis of the human's gaze is also common as an earlier hint to infer the human's desired destination or chosen option. Another common point of these works is that their models usually suffer from uncertainties and error rates that are still not negligible, in spite of the diversity of models tested ranging from primitives analysis and Gaussian Mixture Models (GMM) to more recent Artificial Neural Networks (ANN) in all their varieties. The reasons given for this behavior range from the fact that the model is not yet perfect or that they have not been trained with sufficient data to the fact that the errors are caused by particular cases and outliers and that they can be minimized with the introduction of more complex architectures [31], while allowing the other agent to indicate their intention explicitly when there is a high uncertainty could simplify the problem.

While the correct understanding of the human's intention is essential for the correct performance of the robot, the opposite case is also relevant. This is where the notion of shared intention arises. This concept is originally studied in psychology. Gilbert [32] defines shared intention as a mutual agreement between two or more people to perform a joint action, and [33] adds to this definition a temporal structure in which the actions of the participants are coordinated and take place over time. Dominey and Warneken [34] applies this concept to the field of robotics but using robotics as a tool for designing experiments with which testing theories from psychology.

In actual robotics works, the concept can be found, although treated more tangentially than directly, in fields such as shared control or shared autonomy. Jain and Argall [35] recognizes that effective HRC in shared autonomy requires reasoning about the intentions of the human partner. In their case, a teleoperated grasping task, they find that inferring the robot's goal is difficult even for humans and that both humans and robots make mistakes when trying to predict their partner's intention. Applied to the field of robotic prostheses, in [36] they use three sensors to detect the human's intention (mechanomyography signals, camera and IMU) and they cre-

ate a model that combines this information to improve the control of the prosthesis. However, they still have unacceptable error rates ( $> 50\%$ ) in correctly positioning and grasping some objects, which could be improved by allowing the user to have more control when the uncertainty is too high. In [37] they allow sharing control of the prosthesis so that the human takes control when dexterity is needed while the prosthesis is responsible for maintaining control when a robust grasp is needed. The main problem is to know when to switch from one type of control to the other as this change is detected from EMG (electromyography) signals which are considerably noisy. This allocation of more or less control is what [38] calls arbitration, presenting in the article several methods to achieve it, although maintaining the limitation that the robot must infer the intention of the human and without considering that the human can deliver it directly to the robot.

It is also possible to find completely theoretical articles [39, 40] that study this concept of shared intention and its importance to move from an instrumental interaction to a collaboration [39] or its importance together with shared awareness to create interactions that are transparent (reduce the uncertainty about the behavior of an automaton) for the human [40], but without creating any framework beyond stating the relationships between this concept and others or performing any experiment to support it. In contrast, our work does both of these things, in the line of [38], but with the addition of statistical studies that demonstrate its validity and acceptance by the human.

The approach of allowing the human to explicitly express their intention is not so common. Mullen et al. [41] presents a system in which the robot autonomously performs a manipulation task and, when faced with high uncertainty in meeting any of its goals, informs the human of the various possibilities and asks for their explicit help. Although this work explicitly takes advantage of the human's capabilities to avoid errors, it is oriented to perform autonomous and non-collaborative tasks. Che et al. [42] makes a mobile robot navigating in the presence of humans to indicate its intention both implicitly (by making movements that are legible) and explicitly (by alerting all nearby humans of its presence through a vibration in a wristband on each human's wrist). However, this work takes the opposite approach to ours, i.e., it is the robot that indicates its intention for the human to take it into account rather than the human informing the robot so that it can refine its planning. In any case, they also conclude that using both types of communication increases trust in the robot. For its part, Gildert [43] makes use of this idea to improve object manipulation between two robots by communicating their plans implicitly through the force exchanged through the manipulated object and explicitly by exchanging wireless messages in a pre-established code. With these works in mind, ours differs from Gildert's in that we take the idea of allowing the other agent to communicate explic-

itly to human–robot collaborations rather than robot–robot cases. It also differs from Mullen's in that we orient to collaborative tasks that require repeated and near-constant human–robot interaction in order to be accomplished rather than to tasks performed autonomously by the robot. Finally, our work also differs from Che's in that it is the human that can express themselves both implicitly and explicitly rather than the robot.

The work most similar to ours is [44]. In it they perform a human–robot collaborative search task and also obtain the explicit intention of the human, i.e., what area they intend to explore next, using a mobile application as an interface [45]. However, they neither integrate both types of intention (the implicit intention could be obtained by analyzing the trajectory described by the human) nor present a theoretical framework which can be used in other use cases. In our case, we do present a theoretical framework that allows the integration of both types of intentions and that is general enough to be applied to tasks as disparate as collaborative search or collaborative transportation, which is the main use case used in this article. In any case, more theoretical articles [46, 47] recognize the utility of combining both types of intentions and point out that there are no current system or implementation which makes full usage of that.

Situation Awareness can be useful to integrate both intentions and the environmental information. This concept is, according to the author [6, 48], the understanding of what is going on around you. That is, to keep all the information that is important to the task at hand and discard the irrelevant data. It can be understood through three incremental levels [49]: the first one includes the acquisition of surrounding information. The second one takes care of this different information sources and integrates them considering their relevance. Finally, the third one uses the comprehension of the current state to make future predictions. This has been considered as a key factor for proper decision-making in air combat environments, since it was originally designed for the aviation sector. However, the robotics field has mainly taken advantage of it to design user interfaces [50–57] that include as much information as possible without being obtrusive for teleoperation tasks [58–60] with the objective of increasing the Situation Awareness of the human but not the one of the robot or the set of robots controlled by the human. Other works [61] start from this concept but reformulate it or reorganize its parts adjusting it to the needs of their task. It is worth mentioning the work of [62], in which they use the Situation Awareness concept both to improve the user interface and to process sensor data to allow a Task Reasoner to choose the action to be executed. However, they apply it directly to the specific task of robotic surgery while we use it in a general framework that can include multiple methods of information processing and adapt to multiple tasks, even though we choose afterwards a task as an example.

Focusing on object transportation tasks, [8–10] are known examples of improving the robot’s adaptability to human’s preferences, but always on the basis that the robot is a perfect slave or follower of the human, who acts as master or leader. Other works try to detect human’s role by allowing a switching in the robot’s role among leader, follower and even collaborative [7, 30, 63]. The relation between the measured intent and this role allocation is explored in [38]. In that work, they consider that the interpretation of the human’s intent done by the robot is the one that allows to establish a shared-control policy. By its part, the concept of shared-control, was also widely studied in the literature [64–66]. If we leave aside the more theoretical nature of this article and focus on the use case discussed, collaborative transport of objects, our work also seeks that the robot can adapt to the intention of the human as in [8, 10, 38]. But, unlike these, our work takes into account more roles for the human such as the neutral or the adversarial role. This allows us to explain situations in which the human is behaving in a way that goes against the correct development of the task, a case usually ignored in the literature but not infrequent. Focusing on more technical aspects, our system is based on the one designed in [63] but with the advantage that we only need one force sensor instead of the two used in their implementation.

Finally, [14] was originally considered to model the movement of pedestrians in crowded spaces but it has served as a basis to represent the environment in which a robot should move, specially in urban areas where the robot should share spaces with humans most of the time. [67–69] are examples of trying to make the robot navigate in urban areas avoiding collisions in a socially-acceptable way. The 3D version has also been studied with [70, 71] being examples of implementations of this model adapted to aerial robots. Finally, the previously mentioned [44] is an example of this model used to perform a collaborative task and not an autonomous navigation. We have sought inspiration in this work as well as in [63] to build our force-based model.

In summary, our work differs from others in the literature in that it provides a generic theoretical framework that can be applied to multiple tasks rather than being limited to the task analyzed in each article. This will be shown in Section 3. In addition, it contemplates more roles for the human than the typically considered taking into account the possibility that the human is opposed to the task and does not always collaborate with the robot, either intentionally or unintentionally. This is presented in Section 3.4. Subsequently, we choose a use case such as collaborative transport of objects and design a simple but working model to allow the robot to carry out the task by solving or simplifying some technical aspects present in other works also focused on this use case. This model is presented in Section 4

### 3 Perception–Intention–Action Cycle from Both Perspectives

Think about the previous task, autonomous urban navigation. As a first approach, a robot can detect its environment, processes all the possible obstacles and plan through them. A more elaborated system will consider the humans present in the area as moving obstacles and an even more sophisticated one will calculate their velocity and acceleration to make an estimation of their future movement with increasing uncertainty over time, bigger or smaller depending on the approach [67, 69]. However, if the robot knew where each one of the present humans intend to go, robot’s calculations would be much simpler and the uncertainty considerably smaller. Similarly, humans could reduce the mental burden of having a foreign agent navigating among them if they were aware of its intention as shown in [42].

This idea is what gives rise to our entire theoretical approach: to create a cycle that includes the intention, both implicit and explicit, of the agents collaborating to achieve a better understanding of the current situation, which we consider to be key to obtain both anticipatory and proactive behaviors. Our cycle is shown graphically in Fig. 2. In the following subsections we will explain its different parts.

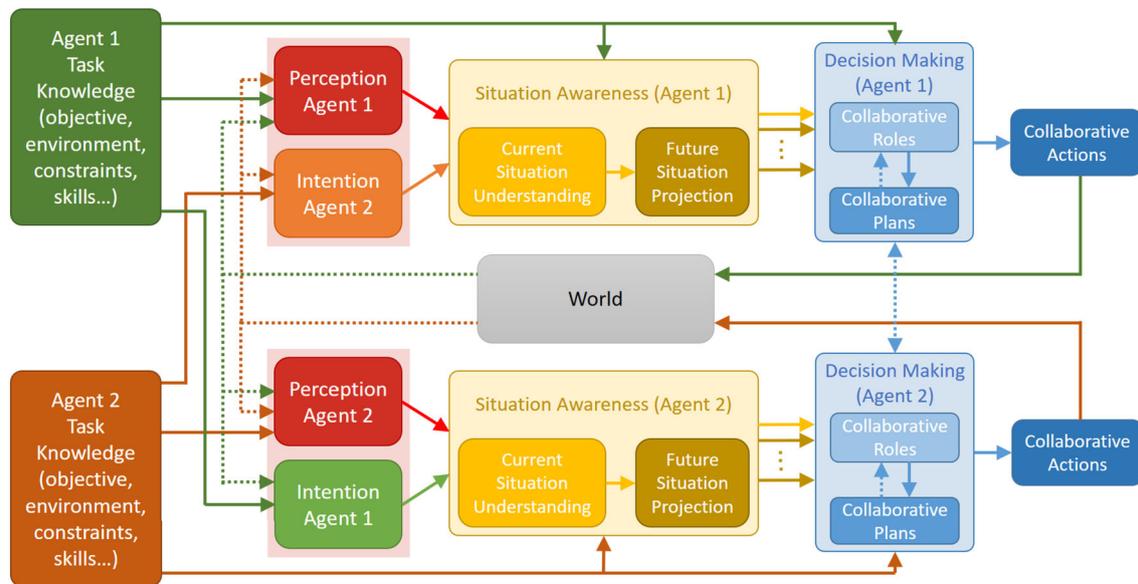
#### 3.1 Task Knowledge

The task knowledge block includes all the previous knowledge that each agent has about the task to be executed. This knowledge does not necessarily have to be the same among all the agents involved. This block can include the objective of the task, the environment in which it is going to be developed in case it is previously known or possible limitations as well as skills to execute the task of each agent. This knowledge is totally task-dependent, so it must be formulated for each specific case.

Applied to the field of robotics, this prior knowledge about the task can be formulated mathematically in each task or can be learned using a model-free approach based on Reinforcement Learning [72–75]. Recent work in this regard is promising by generating a model-free control law to make a biped robot to walk by learning the constraints of the task it is executing [74], or by learning the proximity limit accepted by the human that the robot can approach to and which is dependent on the task that the human is executing [75]. There is also the possibility of modeling this information by combining model-based and model-free solutions [76] or transitioning from one type to the other during the learning process [77].

#### 3.2 Implicit and Explicit Intention

As commented in previous sections, human’s intention is not always inferable. Imagine for example two people moving



**Fig. 2** General information flow according to the Perception-Intention-Action cycle from both agents' point of view in a collaborative task. Previous knowledge about the task is used by both agents. Agent 1 uses this information to perceive their environment and to infer other agent's intention. Agent 2 uses their own information to express their intention explicitly when necessary to avoid misunder-

standings. The situation awareness comprehends the current situation and make projection(s) into the future. This projection allows to establish a collaborative plan according to the role each agent is showing at the moment. This plan generates the following actions which can be perceived again initiating a new cycle. The other agent executes the same cycle

together a bulky object, i.e. a long table or any piece of furniture, in a side-by-side configuration. When one of them (e.g. agent 1) perceives through the force exchanged through the transported object that the other one (e.g. agent 2) starts to turn, they do not know whether they are doing so because they are going to make a turn or because agent 2 wants to move in front of/behind agent 1 to pass through a narrow corridor. Therefore, from agent 1's point of view, agent 2's intention is not clear. Imagine now that agent 1 is carrying the object behind agent 2 so that they have partial information of the environment due to occlusions by the object itself. If it is necessary to avoid an obstacle on their path or simply to stop because they have already reached their destination, the agent in front will not stop abruptly or it could cause an accident. In both cases, what agent 2 will do is to inform their partner. "We stop here", "watch out for the step" or "we change configuration to fit through here" would be the usual phrases in order to eliminate uncertainties or simply to fulfill the task at hand.

This behavior, common between two humans, becomes even more necessary when one of the partners is a robot. This is due to the limitations in the robot's sensors or computational capacity, forcing to use compressed forms of representing perceived information (occupancy maps, segmentations, etc.) that may differ from that used by the human [78], favoring the appearance of errors and misunder-

standings that can be minimized if the human can directly communicate its intentions.

Let us think of a more complex example: a long pass in a soccer game. The player carrying the ball looks at the position of their teammates and opponents and tries to infer their intention (which of their teammates is going to start running towards the opponent's goal and which is not). Based on this information they choose and execute their pass. If they make a mistake, they lose the chance to score a goal, while if they are right, they could make an assist. To reduce the error rate, as well as the mental pressure, it is common to see in professional teams that players nod or shake their heads or signal in some disguised way where they intend to move. In other words, they explicitly state their intention.

This is why it is so important to take explicit intention into account and why in our cycle we separate the classic "Perception" block into two blocks: a "Perception" block in charge of perceiving the environment and another "Intention" block in charge of capturing both types of intentions. This does not mean that no perception is required in the "Intention" block, but that its purpose is different. While the "Perception" block is in charge of perceiving the information necessary to understand the environment, the "Intention" block is in charge of receiving the information necessary to understand the human. This includes perceiving both their actions in order to deduce the implicit intention and the explicit intention itself (the human's voice saying what they want, the human's gesture

indicating what object they want) which is directly delivered so it is only necessary to know the common communication code.

This division allows us to return the Perception stage to its original task of perceiving the environment as well as all the tasks associated with it (SLAM [79, 80], autonomous vehicles [81]...) and to differentiate it from everything related to tasks in which it must collaborate with a human. This can minimize errors, misunderstandings and problems related to having partial information [44], in addition to facilitate the development of algorithms, making them more modular.

Note that it is the correct understanding of the intention of our peers that allows us humans (and, consequently, robots) to act proactively, i.e. not only adapting to each other's actions, but proposing a better plan if we are acting suboptimally. It is worth mentioning that by this we do not mean that we should dispense with implicit inferable intention and rely solely on explicit intention, but that both should be taken into account and processed together to improve performance in any collaborative task since both are contained in the cycle.

### 3.3 Situation Awareness

In order to jointly process both the information from the environment and the intention (both implicit and explicit) of our partners, Situation Awareness comes into play. This concept was presented by Endsley and Garland in [6] and it explains the mental process that allows us to know and understand what is really happening around oneself. It implies using the information received and the previous own knowledge to understand the current situation sifting all the irrelevant data out. For example, if you are looking for a child and you see a group of twenty people, you will rule out anyone taller than a child and you will not waste time looking at their faces. Once this is done, a projection of the future possible situation(s) can be done.

Returning to the previous example with two people moving a heavy object, agent 1 perceives that there are two possible routes, one wider and the other narrower, and hears their partner explicitly telling them that they are going to stand in front, otherwise they will not fit. With this information, agent 1 comprehends the movement that their partner is starting and projects that they are going to go through the narrow path so they can start collaborating towards this goal.

If we go to other sportive example, in the case of an American football match, knowing the position of their teammates and opponents, as well as their intention, is what allows the quarterback to choose which play is the most likely to be successful or even change the play on the fly if they find that the situation has changed.

Applied to robotics, the power of this concept lies in the fact that it is able to organize multiple works that have been proposed in the literature and bring them together under the

same umbrella, allowing, on the one hand, to include them under a common term and, on the other hand, allowing other researchers to recognize which stage of the Situation Awareness they are working on. Works on issues of perception and sensor fusion [82–86] would correspond to the first stage, while works on modeling human actions to try to understand their goal [3, 26, 87] could be included in the second stage. Finally, articles on predicting human movement or future action [4, 28, 67, 88] could be placed in the third stage.

Thus, in our cycle in Fig. 2, the SA is in charge of processing the information received from both the "Perception" and the "Intention" blocks in order to understand the situation in which the human–robot pair finds itself and to predict possible future situations.

### 3.4 Collaborative Task Roles

Once we understand what our partners are doing and we can predict what they are going to do, we can assign a role to each agent based on the task. We consider the classical roles of *master/leader*, *slave/follower* and *collaborative* common in the literature, but we also seek inspiration in other works [38, 89, 90] to consider the *neutral* and the *adversary* role.

For the sake of clarity, we define the *collaborative* role as the one exposed by partners who consider each other as equal peers and who contribute with their knowledge and skills (not necessarily equal among them) to accomplish the task. Likewise, we assign the *leader* role to the agent who imposes their ideas about how the task should be performed and the *follower* role to the agent who accepts this vision and fulfills the *leader's* plans. It is worth mentioning that these three roles carry with them the implicit objective of fulfilling the task satisfactorily. The same is not true for the other two. We consider an agent as *neutral* if they neither act in favour of the task nor against its correct performance and as an *adversary* if their intention is manifestly against accomplishing our task.

In the example of the bulky object transportation, both agents act collaboratively as they both provide force to move the object. As soon as one of them takes the initiative to move in front/behind in order to pass through a narrow corridor, this agent acts as a leader and the other as a follower if they do not oppose to this movement. Both cases of competitive sports serve to illustrate the adversarial role. In both cases, for each player on each team their teammates act collaboratively as they are all contributing with their skills (different among them) to accomplish the task. Likewise, the players of the opposing team act as adversaries as they will act to avoid the previous task.

Applied to robotics, the leader and follower role as well as the collaborative one have been extensively studied [7, 30, 63] but not the neutral or adversarial cases, which does not mean that they do not occur. The adversarial role can be

assigned in situations where robots are used in the presence of children who prefer to play with the robot, making its movements or, in general, the task that the robot was to perform impossible. It also happens with people who have never worked with a robot before and, therefore, do not know the correct way to perform the task with it. In both cases, detecting and assigning an adversarial role to the human allows to an automatic change of strategy aborting the task in case of being in the presence of a curious child or try to communicate with the human partner if it is detected that it does not know how to execute the task. Similarly, the neutral role can be assigned to pedestrians who are unaware of the robot's presence in urban navigation tasks (the collaborative role could be assigned to a passer-by slowing down or braking to let the robot pass).

### 3.5 Action Planning and Execution

As shown in our cycle in Fig. 2, once the SA reports what is happening and what can happen, the sequence of actions to be executed can be selected [91, 92]. We consider that the first thing to do is to assign a role to each of the agents involved, since this allows to automatically eliminate some possibilities before planning. Considering the previous example, it does not make sense to plan a new movement while a child is playing with the robot arm, while it does if the child is moving away and, therefore, exercising a collaborative role. Once a role has been assigned to each agent, it is possible to plan both the actions to be performed by the agent executing the cycle and the actions that the other agent is expected to perform. It should be noted that this planning can include a change in the roles exercised. This is useful if what is intended is to be proactive, that is, to perform an action that causes a change in the other agent in order to convert an adversary into a collaborator or a neutral agent that stops interfering in the task. This assignment-planning process can be repeated for each of the projections delivered by the SA or only with the most probable one depending on the temporal requirements of the task or the agent's computational capacity. Once planned, each agent can execute their actions causing a change in the environment that may or may not be perceived by both agents initiating a new cycle.

This process is executed simultaneously by both agents. In the case of the human, using the innate or learnt through experience cognitive processes and, in the case of the robot, using modules of perception, processing, prediction, planning and action present in the literature.

Two last details should be mentioned. First, that we have not added an explicit stage of agreement between the two agents. This is because such agreement can occur both implicitly and explicitly. For example, in the case of the transport of the bulky object, when one of the agents explicitly expresses that they wish to position themselves in

front/behind to pass through a narrow passageway, the other agent will accept this proposal by facilitating the maneuver (or simply not resisting) and will reject it by exerting force against it or by explicitly expressing their intention to refuse. This may be perceived by both agents initiating a new cycle until they reach an agreement. The alternative is that both agents decide to pause momentarily, share their plans and come to an agreement before resuming the march. This possible exchange of information is what represents the dashed arrow between the two stages of decision making. This possibility is explored in [44], in which the human must explicitly accept or reject the planning done by the robot for both agents.

Secondly, not all the stages of this cycle must always be executed because some of them would not present significant differences with the previous execution. In other words, from a conceptual point of view, it can be understood that this cycle is executed at a variable speed: slower when new relevant information requires making a new prediction completely different from the previous ones, forcing a new decision making process, and faster when the last prediction is being fulfilled and the decision making/planning process can be bypassed. From a technical perspective, the cycle can run at a constant speed but with the high-level modules with low sensitivity to minor variations in their inputs.

This theoretical framework will be tested in Section 4 applied to a specific task such as collaborative transportation. The task knowledge will be presented as constraints and both the environment and the implicit and explicit intention will be modeled using a force model in Section 4.1. Subsequently, this model will be experimentally tested in Section 5.

## 4 Collaborative Transportation as Use-Case

As mentioned in the introduction, we will use a collaborative lightweight rigid object transportation task as a use case to test our proposal. Using this task, we will try to emulate several of the discussed situations with the example of two humans transporting a bulky object. Our goal is to verify with a first round of experiments that the human accepts to indicate their intention explicitly understanding that this can reduce the probability of error and misunderstandings with the robot and that this improves the performance of the task. In a second round of experiments we will add a force predictor, which will do the job of the third stage of the SA. To achieve this, we need a model that allows us to combine four elements: (1) the human's contribution to the task through the force exerted on the transported object, (2) the robot's perception of the environment, (3) the human's implicit intention that can be inferred from the force exerted, and (4) the human's explicit intention indicated by other means.

Thus, the following subsection will show the force model developed for this purpose. This model will be responsible

for combining the information received by the robot from the environment and from its human companion and combining them appropriately to obtain a resultant force that represents the current situation in which the robot finds itself. Therefore, this would represent the first two stages of the SA, leaving the third for the aforementioned force predictor.

The reason for using a force model and not another approach is because this allows the physical variable associated with human effort to be used directly without the need to translate it into any other format. In addition, the Social Force Model (SFM) [14] serves as an inspiration to represent the environment with virtual forces. This allows us to obtain a model with low computational requirements that is easy to understand and implement. Some implementation details will be indicated below with the intention of increasing the reproducibility of our experiments, as well as facilitating the development to any other researcher who wishes to use our model.

We are aware that this model may be difficult to extrapolate to those tasks in which forces are not exchanged. However, its main purpose is to serve as an example to explore several of the concepts presented. It should be stated that, in this first exploratory work, this force model will not be used to provide the robot with proactive behaviors.

Regarding the prior task knowledge, we will assume the following constraints:

- The map will be known to both agents and it consists of a set of valid positions collected in a set  $M \in \mathbb{R}^2$ .
- There will be a set of obstacles:  $O \in \mathbb{R}^2$ ,  $O \subset M$ .
- The goal of the task will be a valid position within the map known to both agents:  $goal \in \mathbb{R}^2$ ,  $(goal \in M) \wedge (goal \notin O)$ .
- The speed of the robot will be limited:  $v_{robot} \in \mathbb{R}^2$ ,  $v_{robot} = [v_{robot_x}, v_{robot_y}]$  with  $v_{robot_x} < v_{lin}^{max}$  and  $v_{robot_y} < v_{ang}^{max}$ .
- The number of involved agents will be  $N = 2$ , being the robot the agent 1 and the human the agent 2.
- The environment is represented by a force  $F_{E,C} \in \mathbb{R}^2$  parallel to the map  $M$ .
- The force exerted by the human  $f_{human} \in \mathbb{R}^2$  is parallel to the map  $M$ , being projected on this plane in other case.

We will use a global planner to calculate the robot's original plan as a succession of waypoints or partial goals to reach the task's goal, in this case, the place to locate the object. If they were two robots collaborating, a shared planner would be enough but, since there is a human in-the-loop, we will try to understand their actions to condition and select the final actions to be performed by the robot.

For that, we will use a force sensor attached to the robot's wrist which is in contact with the transported object. In this way, the human exerts a force on the other end of the trans-

ported object and this force propagates through the object to the wrist of the robot where the force sensor that measures it is located. To simplify the control of the robot, this measured force is projected onto the  $xy$  plane eliminating the  $z$  component that would not produce any movement. In turn, the sensor can measure forces up to 540 N on each of its axes. However, we decided to saturate its measurement to 12 N so that the human does not need to exert great effort to perform the task and so that there is no reward for exerting excessive force that could compromise the integrity of the sensor. Details about how to transform the force measured at the sensor into the force actually exerted by the human will be discussed in the following subsections.

#### 4.1 Perception–Intention–Action Force-Based Model

We use [63] as a starting point. In that work, a lightweight rigid object is also transported collaboratively among two agents. Nevertheless, they do not model the environment by any means. So, as they also did, we will attach a frame  $C$  to an arbitrary point over the transported object and assume that the dynamics of the object in that frame are determined by the joint action of two forces: the one performed by the robot at one end of the object and the force exerted by the human at the other end. In turn, the force exerted by the robot is determined by its interpretation of the environment (present obstacles, location of the task's goal...). Therefore, we can define the task force to be exerted on frame  $C$  as:

$$F_{Task,C} = F_{E,C} + F_{H,C} \quad (1)$$

being  $F_{E,C}$  the component due to the environment and  $F_{H,C}$  the component due to the force exerted by the human.

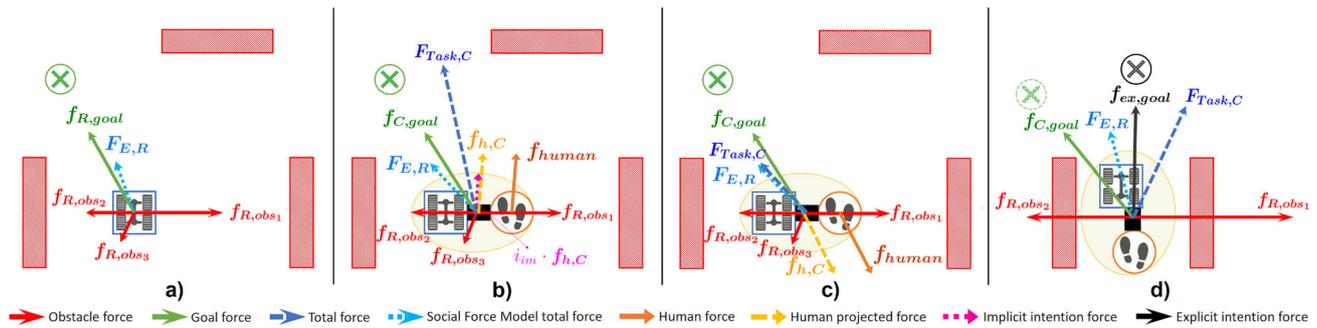
##### 4.1.1 Environment Perception

In order to model the task's environment using virtual forces, we can start from [14]. According to this work, the accelerations and decelerations of a passer-by walking along a crowded street are determined by the joint action of virtual repulsive and attractive forces according to the following expression:

$$f_{a,o} = U_{a,o} (\|r_{a,o}\|) \quad (2)$$

being the virtual repulsive (attractive) force generated by object  $o \in O$  over the agent  $a \in \{r, h\}$ ,  $f_{a,o}$ , the result of applying a monotonic decreasing (increasing) potential  $U_{a,o}$  over the vector  $r_{a,o}$  which goes from the agent to the object.

These objects in  $O$  can be obstacles or the goal. In the first case, we can consider that all the obstacles in the robot's field of view will generate a virtual repulsive force decreasing



**Fig. 3** Response of the Perception-Intention-Action Force-based Model in different situations. **a** Robot navigating alone without any input from the human. **b** Human-robot pair navigating collaboratively with the human contributing with their effort to fulfill the task. **c** Human-

robot pair navigating together but not collaborating since the human is making force to oppose the task. **d** Human-robot pair passing collaboratively through a narrow pass. Forces not normalized for clarity

with the distance from a maximum  $f_{rep,max}$ . If the obstacle is further than a threshold  $d_{max}$  we will not consider it:

$$f_{a,obs} = \begin{cases} f_{rep,max} & \|r_{a,obs}\| < d_{min} \\ 0 & \|r_{a,obs}\| > d_{max} \\ f_{rep,max} \cdot 10^{\alpha \cdot \frac{\|r_{a,obs}\| - d_{min}}{d_{max} - d_{min}}} & \text{otherwise} \end{cases} \quad (3)$$

with  $r_{a,obs}$  the vector from the agent to the nearest border of each obstacle and  $\alpha = -\log_{10}\left(\frac{f_{rep,min}}{f_{rep,max}}\right)$  the decay constant. Likewise, the goal generates a virtual attractive force from  $f_{att,max}$  to 0 with  $d_{goal}$  the distance to start slowing down:

$$f_{a,goal} = \begin{cases} f_{att,max} & \text{if } \|r_{a,goal}\| > d_{goal} \\ f_{att,max} \cdot \frac{\|r_{a,goal}\|}{d_{goal}} & \text{otherwise} \end{cases} \quad (4)$$

In our case, the goal will be each of the successive waypoints of a collision-free route calculated with a global planner so the virtual attractive force will be equal to  $f_{att,max}$  except for the waypoint(s) close enough to the task's goal when this force will start to decline. The way to introduce the constraint related to the task goal is through this global planner: if the goal is not a valid position, the global planner will not generate a valid route so there will be no attractive force. The joint action of all the virtual repulsive forces and the virtual attractive force gives a total force,  $F_{E,C}$ , which represents the effect of the task's environment calculated at the frame  $C$  using  $w_{Rep}$  and  $w_{Att}$  as weights to balance both types of forces. Figure 3a shows a simplification of these forces calculation process for the robot if it were autonomously navigating.

$$F_{E,C} = w_{Rep} \cdot \left( \sum_{obs=1}^{O-1} w_{obs} \cdot f_{C,obs} \right) + w_{Att} \cdot f_{C,goal} \quad (5)$$

Since the total virtual repulsive force's amplitude depends on the number of obstacles and this number can vary from one LiDAR detection to the next one, it is necessary to normalize their addition using  $w_{obs}$  ( $\sum_{obs=1}^{O-1} w_{obs} = 1$ ) weights in order to never exceed  $f_{rep,max}$ . More details about this normalization of the obstacle forces will be shown in Section 4.2.2.

The maximum amplitude for the attractive,  $f_{att,max}$ , and repulsive,  $f_{rep,max}$ , forces can differ, for which we introduced  $w_{Rep}$  and  $w_{Att}$  to obtain a common maximum value,  $f_{max}$ .

$$f_{rep,max} = f_{max} \cdot \frac{1}{w_r}, \quad f_{att,max} = f_{max} \cdot \frac{1}{w_a} \quad (6)$$

$$f_{max} = w_r \cdot f_{rep,max} = w_a \cdot f_{att,max}$$

With  $w_r$  and  $w_a$  the values that equal the maximum repulsive force and the maximum attractive force. Selecting the waypoints which generates each partial goal in such a way that there are no obstacles between the collaborative pair and the following goal, it can be inferred that  $w_{Rep} < w_{rep}$  and  $w_{Att} > w_{att}$  must be fulfilled to ensure that the maximum repulsive force is always smaller than the maximum attractive force to make the robot to move towards the next goal avoiding the known local minima problem associated to the use of the SFM [71]. This ensures that the environment force will tend to reduce the distance to the next goal except if  $\|r_{a,goal}\| < d_{goal}$  when  $F_{E,C}$  could be 0. This implies that the equilibrium point will be reached at a distance  $d < d_{goal}$  closer or further to the goal depending on the number and on how close the obstacles are to it.

Both  $w_{Rep}$  and  $w_{Att}$  can be variable and updated according to some policy that meets the above restrictions. In this case, both weights will take constant and equal values for all experiments as will be indicated in Section 5.1.

Additionally, this environmental force can be normalized to impose  $f_{max}$  as maximum in order to make it comparable

to the measured maximum human’s exerted force if we select  $f_{max} = f_{human,max}$ .

$$F_{E,C}^{norm} = \frac{F_{E,C}}{w_{Rep} + w_{Att}} \tag{7}$$

### 4.1.2 Human Intention

Instead of just adding the human’s exerted force to  $F_{E,C}^{norm}$ , we can use this last force as well as the virtual generated forces to understand the human’s intention, which we postulate that be divided into two terms: implicit intention and explicit intention.

The first one is calculated as follows. The force exerted by the human on the other end of the object,  $f_{human}$ , is detected at the robot’s force sensor,  $f_{sen}$ , as the difference between the force exerted by each agent over the object:

$$f_{sen} = f_{human} - f_{robot} \tag{8}$$

If the robot moves without acceleration, the force due to this agent,  $f_{robot}$ , will be 0. In fact, if both the human and the robot move at the same velocity and the human does not desire to accelerate or slow down, no force will appear at the sensor. In addition, the robot knows its own movement so it can discount its effect as well as the objects weight to calculate the force due to the human’s action at the sensor:

$$f_{h,sen} = R_{h,sen} \times f_{human} \tag{9}$$

being  $R_{h,sen}$  the rotation matrix which eliminates the robot’s force and transforms the human’s force from their grasping point to the robot’s force sensor frame. Likewise, we can transform this force to frame  $C$  using the corresponding rotation matrix,  $R_{sen,C}$ :

$$f_{h,C} = R_{sen,C} \times f_{h,sen} = R_{sen,C} \times R_{h,sen} \times f_{human} \tag{10}$$

Finally, this force is compared with the attractive virtual force previously generated in (4) resulting in a coefficient  $i_{im}$  proportional to the angle between the human’s force translated to the frame  $C$  and the goal force,  $\phi_{h,C;goal}$ :

$$i_{im} = k \cdot \cos(\phi_{h,C;goal}) \tag{11}$$

This angle can be used to detect whether or not the human is cooperating with the task. Thus, if it is below a threshold  $\phi_{collab}$  the robot can consider that the human is collaborating since the force the human is exerting is consistent with the path the robot has calculated to follow. At the same time, if this angle is greater than a second threshold  $\phi_{adver}$ , the human can be considered to be opposing the task since the force they are exerting is opposite to the force that should be

exerted to follow the stipulated path. Finally, if this angle is between the two thresholds,  $\phi_{collab} < |\phi_{h,C;goal}| < \phi_{adver}$ , nothing can be stated about the human’s contribution. In the first case, the robot can assign them a collaborative role; in the second, an adversarial role; and in the third, a neutral role. This way, in (11)  $k = 1$  if it is detected that the human’s force goes in favour of the task (i.e., the robot assigns to the human a collaborative role) and  $k = 0$  if it goes against the task (neutral or adversary role).

The  $\cos(\cdot)$  function can be replaced by any function that is maximum for  $\phi_{h,C;goal} = 0$ . In this case, the  $\cos(\cdot)$  function has been chosen so as not to overly penalize small deviations between the force exerted by the human and the expected one.

As for the explicit intention, this term depends on the sub-task and on the means of communication used by the human. In general, we convert this explicit intention to a change in the environment. The general expression for this term is similar to (2) but taking into the account that this term is valid for a limited period of time  $t$ .

$$f_{ex,o} \propto U_{a,o}(\|r_{a,o}\|, t) \tag{12}$$

We generate a explicit intention vector for each of the commands given by the human. For example, if the human indicates that they wish to avoid a particular path, this command will generate a virtual obstacle which generates an extra repulsive force. If they indicate that they wish to pass through a narrow passage, it will substitute the task’s goal by a temporal subtask’s goal at the other end of the passage generating an attractive force.

### 4.1.3 Situation Awareness

With all the virtual forces and coefficients calculated in the previous steps, it is possible to check if the implicit intention is relevant according to its coefficient and generates the human’s contribution to the task force:

$$F_{H,C} = f_{h,C} + f_{im} = f_{h,C} + i_{im} \cdot f_{h,C} \tag{13}$$

being  $f_{im}$  the implicit intention force when its coefficient is considered as relevant.

We use this intention as a way to potentiate the human’s force, as it can be seen in Fig. 3b. The human’s force is transformed to frame  $C$  and the implicit intention coefficient is calculated and with this, a collaborative role is assigned to the human. The total force potentiates the human’s force avoiding the obstacle on the left while it keeps going towards the goal. Figure 3c instead, shows the case where the human exerts a force that goes against what the robot would expect them to do. This force is also transferred to frame  $C$  and the corresponding coefficient is calculated but, as it goes against

the task, the human is assigned an adversarial role,  $k = 0$  in (11), and this virtual force is not taken into account.

Likewise, the situation awareness converts the received commands into modifications of the attractive (generating subgoals) and repulsive (generating obstacles) forces taking into account their time validity and the previous understanding of the current situation. Figure 3d illustrates the case of a narrow passage through which the human has explicitly expressed their interest in going through.

Thus, the task force to be applied to the object for each explicit intentions is:

$$\begin{aligned} F_{Task,C_e} &= w_E \cdot F_{E,C_e}^{norm} + w_H \cdot F_{H,C} \\ &= w_E \cdot F_{E,C_e}^{norm} + w_H (1 + i_{im}) \cdot f_{h,C} \end{aligned} \quad (14)$$

being  $F_{E,C_e}^{norm}$  the  $e$  modified version of the task's environment force. Like in (7), the task force can also be normalized in order to delimit it to  $f_{max} = f_{human,max}$ .

$$F_{Task,C_e}^{norm} = \frac{F_{Task,C_e}}{w_E + w_H} \quad (15)$$

This process is repeated with each explicit intention  $e \in I_e$  being  $I_e$  the set of considered intentions. Finally, each task force can be used to generate a future projection of how the object will move and its impact on the development of the task.

## 4.2 Implementation Details

The approach followed by the authors to implement the model into a real robot and solve the collaborative transportation task was based on building a reactive control scheme that could make fast decisions using the latest information coming both from the environment and the human. The importance of the reactivity of the robot's response comes from the human's sensitivity towards the actions of the robot, as both agents are linked together through a rigid object.

Three aspects must be taken into account by any researcher wishing to reproduce our experiments. First, how we detect and process environmental obstacles. Second, how we normalize the repulsive forces so that their sum is comparable to the attractive force and that of the human. Third, how we convert the total force into velocity commands.

### 4.2.1 Environmental Obstacles Processing

The detection of obstacles in the environment is performed by LiDAR sensors, being valid for this task both 2D and 3D sensors, since the 3D LiDAR information can be projected over the plane of the robot making it equivalent to a 2D sensor (although with the advantage of being able to detect

obstacles that have been suspended in height above the plane of the sensor).

In the case of using a 3D sensor, the algorithm starts by limiting the height it takes into account to eliminate the beams towards the floor and the ceiling. The resulting beams are then projected onto the sensor plane. Hereafter the processing is the same for both sensors. First, the points that are at a distance greater than a threshold  $d_{max}$  selected for the experiment are eliminated.

As the cloud generated is referred to the sensor frame, it is necessary to convert it into a common frame using the transformation tree of the robot. Having all sensor clouds referred to the same frame, we will proceed to merge them in case the robot has more than one LiDAR sensor into a single data structure which will be processed. The merged cloud is then filtered to remove outliers points using a statistical filter which analyses the k-points neighbourhood of each point, focusing on the distribution of those points. The points exceeding a certain deviation threshold are classified as outliers and removed from the cloud. All the calculations until here are done using the PCL library for C++ [93].

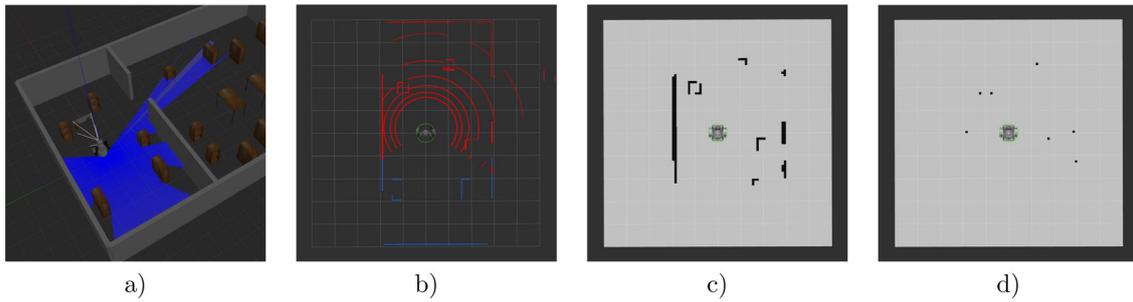
The filtered cloud is projected to the floor plane if it was not already and then used to define an occupancy map to simplify the obstacles detection. As the cloud was previously filtered, we have considered a cell occupied if a single point of the cloud is projected over that cell. The last processing step consist of clustering the obstacles using a connectivity method that uses the 8-neighborhood to make the cells into a higher syntax level: obstacles. From these obstacles, the Social Force Model is only interested in the nearest point to the robot that will be used to compute the repulsive term. In the clustering process it was also established a minimum size for the obstacles to remove small clusters that may be noise.

This processing module has been tested in two social robots such as TIAGo++<sup>1</sup> and IVO [94] with similar results. The first one has a 2D LiDAR in the front part and the second one a 3D LiDAR in the front and a 2D LiDAR in the back, being necessary in the case of the second robot to project the information of the front LiDAR on a plane and later to merge the resulting information with that of the back LiDAR. Figure 4 shows the obstacle detection workflow in a simulated environment using the IVO robot. All the experiments that will be shown and analyzed in the following section were performed with a single robot, TIAGo, to make them all comparable.

### 4.2.2 Normalization of Repulsive Forces

Each obstacle generates a repulsive vector in the direction of the distance vector from the nearest point of the obstacle towards the robot. The module of the vector is bounded by

<sup>1</sup> TIAGo URL: <https://pal-robotics.com/robots/tiago/>.



**Fig. 4** Environmental obstacles detection workflow. **a** Simulation in Gazebo of IVO robot in an indoors environment. **b** Environment detection with front 3D LiDAR (in red) and rear 2D LiDAR (in blue). **c** Occupancy map with the obstacles generated after clusterization of data

from both LiDARs. **d** Occupancy map with the nearest point to the robot of each detected (bigger than a minimum size) obstacle. (Color figure online)

definition, but when there are several obstacles it is not possible to know if the magnitude of the total repulsive term will exceed the maximum set for the total repulsive force.

To solve this problem, one possibility is to directly saturate the magnitude of the vector keeping its direction, but this can lead to saturating the repulsive term in a direction that does not correspond to the one which would be generated by the nearest obstacles. Other possibility is to select a constant number of obstacles following a distance criteria, as the number of obstacles is constant then the term can be bounded directly, or using the magnitude of the nearest obstacle which generates the highest repulsive force [71]. Another possibility relies on experimental data of how humans solve a certain task, from which a fitting process is done for all the parameters of the model until the behaviour matches the data available [95].

However, the first approach misses environment information and the second one is based on experimental data, which can be difficult to obtain for certain tasks and can change between experimental cases. For this reason, we propose a new normalization mechanism of the repulsive term that bounds the total repulsive term without human intervention or prior knowledge of the environment nor of the task.

The normalization procedure computes the mean magnitude of the  $N$  repulsive terms at a given instant to compute a normalization coefficient in (17). Then, the coefficient is applied to each individual term in (18). After that, the total repulsive term is guaranteed to be bounded as can be seen in (19).

$$\bar{R} = \frac{1}{N} \cdot \left( \sum_{i=0}^N \|f_{obs,robot,i}\| \right) \quad (16)$$

$$w_{obs} = \max \left\{ \begin{array}{l} \frac{\bar{R}}{F_{max}} \cdot N \\ 1 \end{array} \right. \quad (17)$$

$$f_{obs,robot,i}^{norm} = \frac{f_{obs,robot,i}}{w_{obs}} \quad (18)$$

$$f_{obs,robot,i}^{norm} = \sum_{i=0}^N \frac{f_{obs,robot,i}}{w_{obs}} \Rightarrow$$

$$\|f_{obs,robot,i}^{norm}\| \leq \frac{1}{w_{obs}} \cdot \sum_{i=0}^N \|f_{obs,robot,i}\| \Rightarrow$$

$$\|f_{obs,robot,i}^{norm}\| \leq \frac{1}{\frac{\bar{R}}{F_{max}} \cdot N} \cdot \sum_{i=0}^N \|f_{obs,robot,i}\| \Rightarrow$$

$$\|f_{obs,robot,i}^{norm}\| \leq \frac{\sum_{i=0}^N \|f_{obs,robot,i}\|}{\frac{1}{N} \cdot (\sum_{i=0}^N \|f_{obs,robot,i}\|) \cdot N} \Rightarrow$$

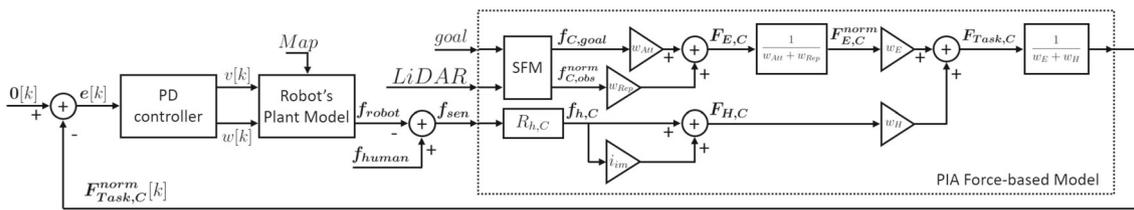
$$\|f_{obs,robot,i}^{norm}\| \leq F_{max} \quad (19)$$

This maximum value  $F_{max}$  can be used to limit both the repulsive and attractive forces. If it is also equal to the maximum force of the human that the system takes into consideration (the human can exert a higher force but it is not taken into account), all forces are comparable. Besides, human forces were bounded to 10 N with a double aim: reduce the necessary effort that a human has to exert to control the robot and protect the sensor and the hand from an excessive effort that may affect the integrity of the robot. The bounded force proved to be comfortable to most humans in the experimental phase.

#### 4.2.3 Robot's Platform Controller

Our approach considers that the robot moves on a plane so the calculated resultant force,  $F_{E,C}^{norm}$ , will have two components,  $x$  and  $y$ , which can be used to calculate the robot's linear and angular velocity. Figure 5 shows a diagram of the control scheme used.

The logic of operation is as follows. While the task is running, there will be a goal and potentially obstacles in the



**Fig. 5** Control structure used to generate the platform's linear and angular velocity. A PD controller is used to generate velocity commands from the task total force. These commands are sent to the robot's internal con-

troller making the robot to do whatever is necessary to fulfill the task taking into account the human's intention

way. This added to the choice of weights in (6), guarantees that  $F_{E,C}$  is not null. At the same time, being a collaborative transport task, the human will have to make an effort making  $F_{H,C}$  non null either, unless the human perfectly follows the robot's movements. This implies that  $F_{Task,C}$  will only be 0 when the task is finished. Therefore, we take 0 as the set-point and  $F_{Task,C}^{norm}$  as the controlled variable. The error is sent to a task controller (in our case, a PD controller) which generates the linear and angular velocity commands necessary for the robot to solve the task by adapting to the human's wishes in the process. This also ensures that, if the human exerts the exact opposite force to cancel the  $F_{E,C}$  component by opposing the task and making  $F_{Task,C} = 0$ , the robot stops. Therefore, the control loop designed follows the common disturbance rejection architecture, in which the environment (obstacles and task) and the human forces are modelled as disturbances that should be rejected to achieved the control goal, being in our case arriving to the goal point of the task. In other words, our objective function to minimize would be:

$$\min_{f_{robot}, f_{C,goal}, f_{C,obs}} \Phi = F_{Task,C}^{norm} \quad (20)$$

where  $F_{Task,C}^{norm}$  is obtained following eq. (15).

It is worth mentioning that this controller is not an action planner, since it is not planning high level actions but generating the necessary velocity commands to move the robot. In this case, with a frequency of 10 Hz as this is the slowest input signal, the LiDAR.

Note that we split the platform control into two blocks: our task controller and the robot's internal wheel controller which would be inside the plant model. This may seem like a weakness of our force-based model, but it is actually a strength as it allows us to decouple our controller from the robot's own controller. This allows us to abstract from the specific robot we are using and use our model in any other robot.

The disadvantage of this is that if the plant model is not known, as it is in our case as we are using the commercial robot TIAGo++ manufactured by the external company PAL Robotics, the stability analysis of the complete sys-

tem becomes significantly more complicated. In our case, we have assumed that the robot controller is stable for the reduced speeds we are considering in this use case. In turn, we have not considered it necessary to perform a stability analysis of the PD controller used since it is a widely known architecture and it was tuned for this specific experiments with very low allowed speeds.

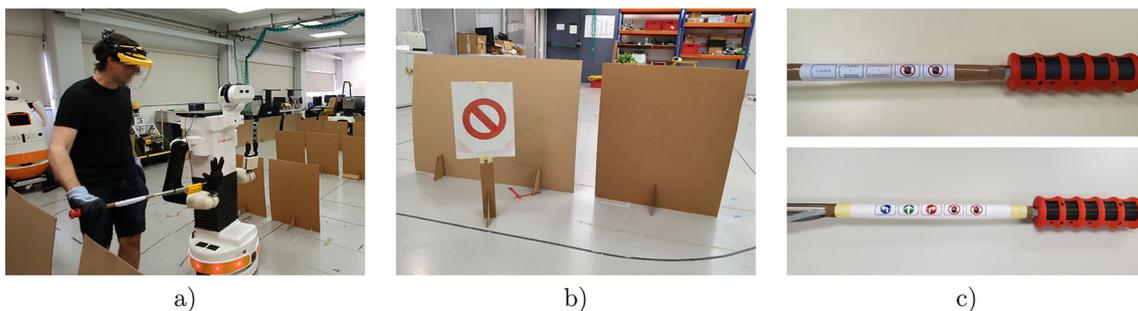
Now that the force model to be used has been presented, two rounds of experiments with this model will be performed in Section 5 to test its validity. Subsequently, in Section 6 some of the limitations of the results obtained will be discussed.

## 5 Experimental Results

The force-based model outlined in the previous section is used to perform two rounds of experiments. With the first round we seek to demonstrate both that humans understand that they should give the robot their intention explicitly when necessary to minimize errors and misunderstandings and that not only it is not bothersome to them but also helps to improve task performance. With the second round of experiments we try to compare our approach (including both types of intention) with a classical system that uses only implicit intention to predict the human's future actions. In doing so, we seek to show the advantages of taking both intentions together into account. The study of proactive behaviors on the part of the robot is therefore outside the scope of this work.

### 5.1 Experiments Setup and Methodology

All the experiments presented in this section have been performed in an indoor environment on a  $7.8 \times 5.7$  m stage with OptiTrack on the ceiling to allow localization of both agents. In all of them a pair consisting of a human and a robot perform a collaborative transport task in which the transported object is an 80 cm steel bar. This task is repeated in different scenarios with multiple obstacles (walls, columns, narrow corridors and/or forbidden passage signs) that limit the possible routes or require a different type of collaboration between



**Fig. 6** Experimental setup. **a** Example of human–robot pair performing one of the experiments. The human wears a protective helmet. Both agents wear OptiTrack markers on the head/helmet to allow them to be precisely located. The scenario has several walls limiting the number of possible routes to the target. **b** Modules used for mounting the walls.

One of the forbidden pass signs used. **c** Handle used for the human to grasp the transported object and to communicate their intention explicitly using the buttons present. Code associated with each button printed next to the handle

both agents in order to avoid them. Figure 6a shows a volunteer performing one of the experiments in one of the designed scenarios. Both the exterior and interior walls are mounted with modular blocks (Fig. 6b) to allow changing from one scenario to the next one as quickly as possible.

As for the robot used, it is a TIAGo++ manufactured by PAL Robotics. It has a force sensor on each of its wrists. On the right wrist of the robot, an accessory is added to rigidly and securely attach the transported object. To allow the human to explicitly indicate their intention, we have designed a handle with 5 buttons, one for each finger, and placed it on one of the ends of the transported object. This improves ergonomics and allows the human to communicate directly with the robot using a robust code known to both parties. The last two buttons (corresponding to the ring and little fingers of the right hand) are not used while the other three represent different commands depending on the round of experiments. This communication system guarantees robustness as long as the only possible failure is due to a human pressing the wrong button. To minimize this, indicative signs are attached to the side of the handle as a reminder so that the human knows at all times the meaning of each button (Fig. 6c). More information on this subject in the following subsections.

Also to ensure the robustness of the experiment, the force sensor is calibrated according to the manufacturer's instructions before starting each round of experiments with a new volunteer to avoid the effects of possible drifts that this sensor may have. For its part, the reliability of the robot's movement is guaranteed by the manufacturer's low-level controller in charge of moving the robot's wheels as long as the speed commands generated by our system are valid.

As for the humans who participated in this study, they were recruited from our research institute, as well as from different schools of the partner university (industrial engineering, telecommunications engineering, mathematics, architecture,

physics and chemistry), which allows us to have a diverse population sample with varied knowledge of robotics from professional experts to no prior knowledge. At the same time, their ages range from 19 to 55 years old, all of them being of legal age and in full use of their mental faculties. No volunteers were paid for participating in this study, ensuring that there is no conflict of interest. A total of 58 people performed 246 experiments, of which the first 135 were in the first round, 39 were used exclusively to train a force predictor, and the last 72 were used in the second round of experiments.

In all of them, anonymous objective and subjective data are obtained. The objective data are obtained from the positioning using OptiTrack, which allows us to obtain measurements of distance traveled and time spent on the task. We also store data from the force sensor to measure the average and maximum force exerted in each experiment. The subjective data are obtained with a hand-made questionnaire that the volunteers fill out in-situ after each run so that they do not forget their impressions. These questionnaires include multiple-choice and rating questions. For the second ones, we use a scale of 1–7, thus giving the volunteer more options to express themselves than with a classic scale of 1–5. All variables analyzed in this article by variance tests are normally distributed according to the Shapiro–Wilk test unless otherwise indicated. Additionally, after finishing the questionnaire, a brief interview with open questions is performed with each volunteer. Appendix A shows an example of the common hand-made questions used in every round of experiments and the structure of the interview.

All experiments have been recorded with a fixed camera in those cases where the volunteer gave consent to be recorded (in addition to their consent to participate in the study and for the rest of the data collected to be used anonymously). Additionally, data relating to the force sensor readings, the robot's LiDAR and the location of both agents using the OptiTrack positioning system have been stored anonymously (assigning

a random numerical ID to each volunteer) for later analysis. In total the experiments take between 30 and 40 min, including fulfilling the questionnaire.

Finally, the weights used are  $w_{Att} = 1.01$ ,  $w_{Rep} = 0.99$ ,  $w_E = w_h = 1.0$ , the thresholds are  $\phi_{collab} = \pm 45^\circ$  and  $\phi_{adver} = \pm 120^\circ$  and the controller gains are  $k_P = -2.0$  and  $k_D = -0.35$  in all experiments. All the experiments reported in this document have been performed under the approval of the ethics committee of the Universitat Politècnica de Catalunya (UPC)<sup>2</sup> in accordance with all the relevant guidelines and regulations (ID: 2021.10) and all volunteers who participated in this study have signed an informed consent form.

## 5.2 Explicit Intention User Study

The main objective of this first round of experiments is to check that the human willingly accepts to state their intention explicitly and that this does not place an excessive burden on them. Having done this, the effects that the inclusion of explicit intention has on the task are tested.

For this purpose, five scenarios are developed, which can be seen in Fig. 7 (the sixth shown scenario will be analyzed in the following subsection corresponding to the second round of experiments); and 27 volunteers (age:  $\mu = 28.29$ ,  $\sigma = 6.58$ ) perform a first round of 135 experiments (5 each). After each experiment, i.e., after performing the task in each of the five scenarios, the volunteer fills in a section of the questionnaire used to assess subjectively various parameters of the model or the task.

The first two scenarios serve as an introduction to let the human know how well the robot can navigate on its own and how the human can control it. Thus, in the first, the robot acts as a leader by making  $i_{im} = 0$  in (14) for the whole experiment to take control of the task. This implies that the human is still able to collaborate with the robot but will have to exert much greater force if they intend to prevail. In the second scenario, the robot becomes a follower of the human by making  $i_{im} = 0$  for the whole experiment and reducing  $w_{Att} = 0.1$  to thus maintain obstacle avoidance but following the human's commands. In this way, the human discovers the robot's response speed and learns to communicate implicitly through their force. As can be seen in Fig. 7a, b, in both scenarios there are two routes to accomplish the task.

The next three experiments confront the human–robot pair with different situations that they must solve to complete the task. In the third, the human detects a forbidden pass sign on the route that the robot cannot "see" (the robot has no system that allows it to interpret the meaning of this sign). The human must use their force to make the robot re-plan once they have brought it to a point where the other route is

shorter. The fourth experiment is similar to the third, but the human can use the buttons on the handle (see Fig. 6c). These include a button to take control of the task, i.e., explicitly tell the robot that the human wishes to become the leader with a consequent increase in  $w_H$ ; and another to command the robot not to continue along the current path, in which case, the robot uses this message to generate a virtual obstacle along the path it is following causing the attractive force that was driving it along that path to become repulsive and force the robot to replan. As in the previous two experiments, in both the third and fourth experiments there are also two routes to complete the task. Since these two routes are the same size, the human can choose the route they want, making it impossible to know a priori which route should be blocked to force the human to go back and take the alternative route, which is the situation we want to force the human to confront in order to check if they understand the utility and necessity of explicit communication. In consequence, two forbidden pass signs were placed, one in each route, so that the volunteer is forced to deal with the situation regardless of the choice. Once a route is picked, a research assistant removes the remaining sign to enable the alternative path. This is done so that the experiments are systematic and do not depend on chance ensuring that all volunteers are forced to deal with this situation regardless of their choices. The human is unaware of the fact that both paths are blocked at the beginning of the experiment.

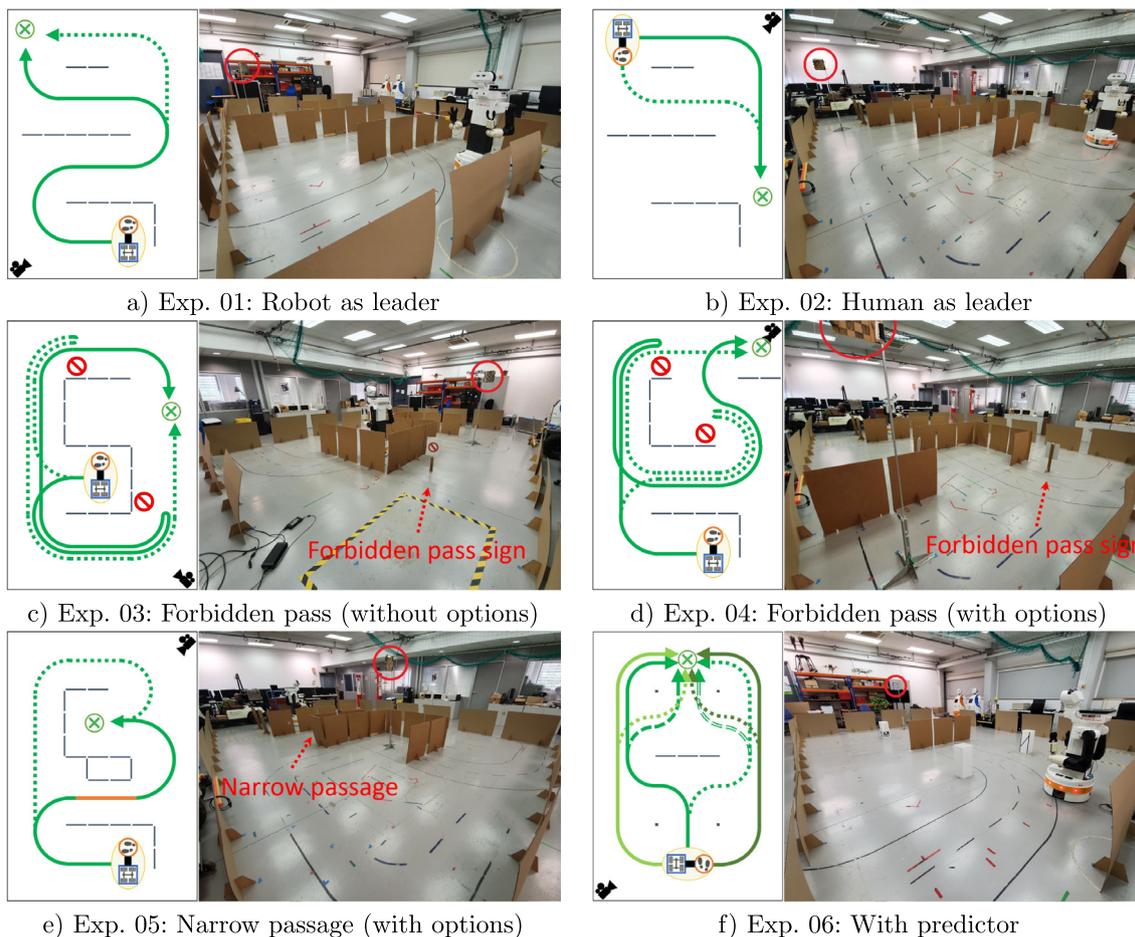
The fifth and final experiment features an overly narrow corridor on the shortest route that forces the human to stand behind the robot instead of beside it in order to pass. To do this, they are allowed to use a third button that serves to communicate to the robot their explicit intention to change the pair configuration by placing the human behind the robot.<sup>3</sup>

The first two scenarios serve for the human to get practice with the robot in the task to be performed and for us to know if the force-based model is comfortable for them or should be improved in following works. Therefore, the answers given by the volunteers to the first two sections of the questionnaire corresponding to the first two experiments are analyzed qualitatively. Being the scale used from 1 to 7 we can use 4 as the threshold that determines whether the human agrees or disagrees with what is being asked.

Thus, Figs. 8 and 9 show that the human considers that they feel comfortable both accompanying the robot when it acts as a leader and guiding the robot when it is the human who leads the interaction. At the same time, they consider that the solutions proposed by the robot are acceptable without being aggressive in its movements. Likewise, they consider that the robot's response speed is adequate and that the robot's control system is sufficiently intuitive. Finally, trust in the

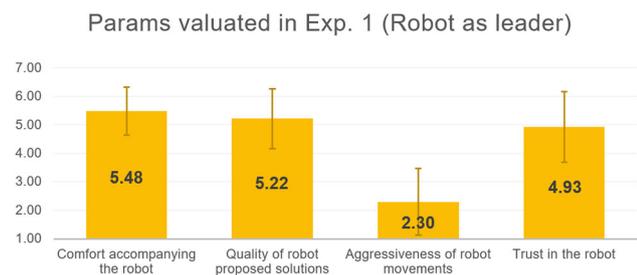
<sup>2</sup> Ethics committee URL: <https://comite-etica.upc.edu/en>.

<sup>3</sup> 1st round example: <https://youtu.be/LO6zLUbnKMU>.

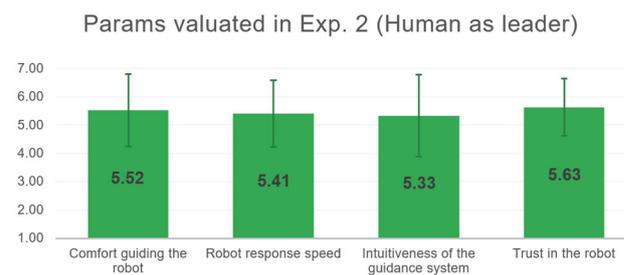


**Fig. 7** Setups for each experiment. The robot is located at the starting point in each experiment. The goal is marked with a green X in the diagrams and with a chequered flag in the real setups. There are two routes in each experiment to fulfill the task except in Exp. 05, where

the human has to change the footprint to enable the second one, and in Exp. 06, where there are at least eight feasible routes. Camera icon on the left represents picture’s point of view on the right. (Color figure online)



**Fig. 8** Valuation performed by volunteers after Exp. 1. Average and std. dev. of the parameters asked to valuate from 1 (very low) to 7 (very high). Error bars represent std. dev



**Fig. 9** Valuation performed by volunteers after Exp. 2. Average and std. dev. of the parameters asked to valuate from 1 (very low) to 7 (very high). Error bars represent std. dev

robot increases when it is the human who takes control of the task. This aspect will be discussed in more depth below.

We can therefore proceed to use this force-based model to test our hypotheses about allowing humans to express their intention explicitly versus not allowing them to do so:

**H1** - Allowing humans to explicitly state their intention allows for a reduction in their effort.

**H2** - Allowing humans to explicitly state their intention increases parameters of effective HRI including perceived safety and trust in the robot.

**Table 1** Subjective (from 1 to 7) and objective (in Newtons) measuring of needed force

Measurement	Exp. 3	Exp. 4	<i>p</i> value
Necessary Force	5.11 (0.93)	<b>3.15</b> (1.26)	< 0.001
Difficulty to comm. intent	3.93 (1.66)	<b>5.56</b> (1.05)	0.005
Avg. Force [N]	8.62 (3.35)	<b>5.73</b> (2.56)	< 0.001
Max. Force [N]	26.1 (7.71)	<b>19.8</b> (7.09)	0.004
Avg. Force last 25% [N]	7.93 (1.54)	<b>5.91</b> (1.43)	< 0.001

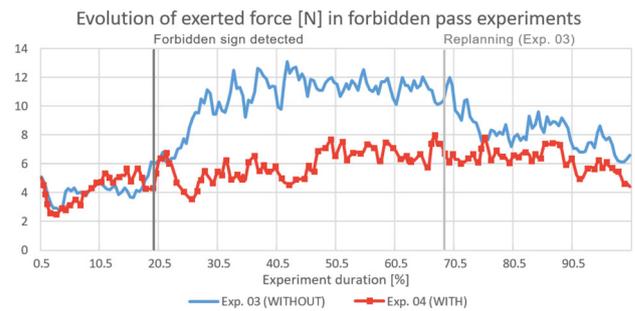
**H3** - Humans understand that their explicit intention is needed to improve task performance.

Before presenting the results, it is interesting to perform a post-hoc statistical power test to know what values we can be statistically sure of, taking into account the sample size used. Thus, using the criterion of  $p < 0.05$  and having 27 volunteers, we can detect effect sizes as low as  $\eta^2 = 0.135$  with a statistical power of 80%.

To test hypothesis H1, we will compare the third and fourth experiments, since both pose the same situation with the only difference being whether or not the human can explicitly indicate their intention using the buttons on the handle. At the end of each of these experiments, in the corresponding section of the questionnaire, the volunteers are asked about the force they subjectively consider they had to exert to make the robot go the other way from 1 (negligible) to 7 (extreme). They are also asked how difficult it was for them to communicate their intention to the robot from 1 (impossible) to 7 (no difficulty). At the same time, both the average force and the maximum force measured by the force sensor in each experiment are analyzed. The results are shown in Table 1. It also shows the *p*-value calculated for each variable using ANOVA tests on the samples of both experiments.

As it can be seen, there is a statistically significant reduction in both subjective variables, taking into account a criterion of  $p < 0.05$ . That is, the human considers that they must exert significantly less force to make the robot change its route when they can explicitly communicate with it ( $F = 42.2$ ,  $df = 53$ ,  $\eta^2 = 0.448$ ), which has repercussions in that they also consider that it is significantly easier for them to communicate their intention ( $F = 18.5$ ,  $df = 53$ ,  $\eta^2 = 0.263$ ). This is corroborated when analyzing the two objective variables, also producing a statistically significant reduction in both the average force and the maximum force exerted.

If the evolution of the force exerted in each experiment is analyzed, a noteworthy phenomenon appears. Figure 10 is obtained by resampling the sequence of force measurements in the third and fourth experiments to make them all have the same duration (same number of samples) so that the samples



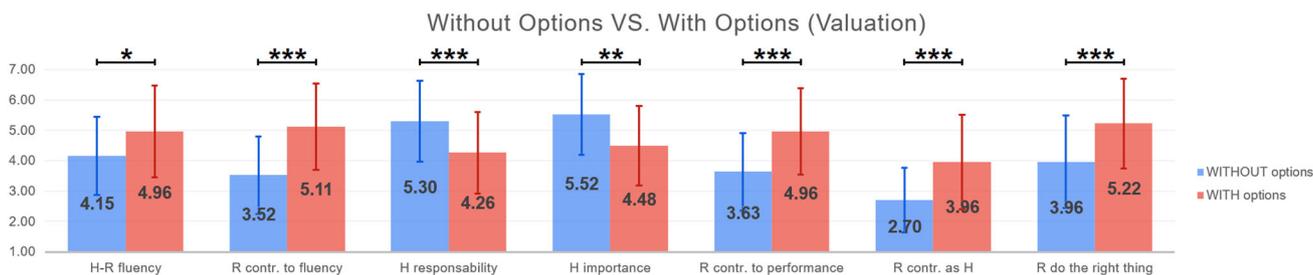
**Fig. 10** Evolution of the average force exerted in Exp. 3 and 4. Extra force needed in the third experiment once the forbidden path sign is seen to make the robot to go backwards until it replans using other route. No extra force needed in the fourth experiment

in each percentage of the experiment can be averaged. As it can be seen, the human must significantly increase their effort to force the robot to go the other route in the third experiment by lacking a way to tell the robot that it has partial information about the task. This extra effort can be avoided in the fourth experiment.

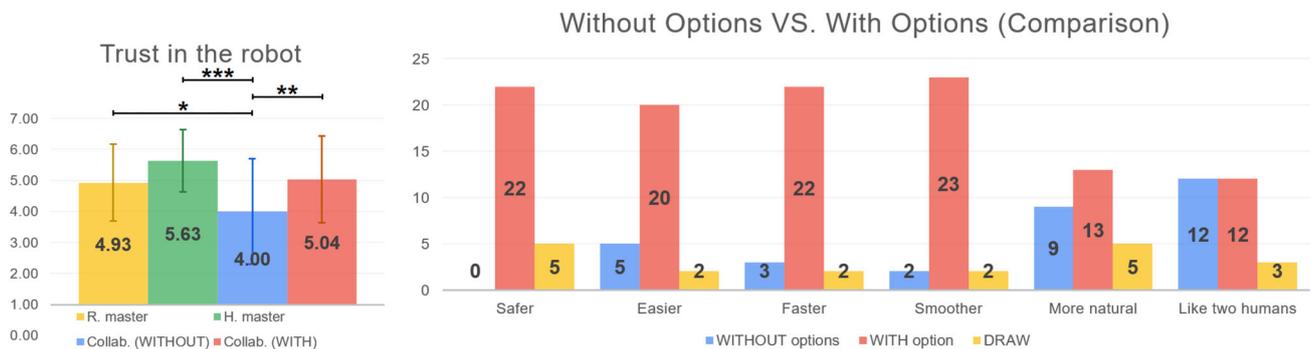
This may seem an obvious result considering the two setups. However, the interesting part is that the user still exerts a greater effort once the robot has replanned, when there is no longer a difference between the two experiments. The last row of Table 1 shows the result of comparing the average force exerted by the human during the last 25% of each experiment. As it can be seen, there is a statistically significant reduction in the effort exerted ( $F = 25.1$ ,  $df = 53$ ,  $\eta^2 = 0.326$ ). Therefore, we can affirm that hypothesis H1 is correct.

In the section of the questionnaire corresponding to the third experiment, volunteers are asked to evaluate different parameters related to human–robot interaction. Volunteers are asked again to evaluate the same parameters after the fifth experiment by evaluating the fourth and fifth together. In this way, we can compare their experience to test hypothesis H2 using an ANOVA test. The results are shown in Fig. 11.

Comparing the fluency of the interaction, there is a slightly significant increase in the perceived fluency of the human–robot pair: without options  $\mu = 4.15$ ,  $\sigma = 1.29$ ; with options  $\mu = 4.96$ ,  $\sigma = 1.50$ ;  $F = 4.40$ ,  $df = 53$ ,  $\eta^2 = 0.080$ ,  $t(26) = -2.16$ ,  $p = 0.0403$ . However, since its effect size is lower than the minimum considered to have an acceptable statistical power, we cannot be totally sure of this result. The same does not occur if we look at the robot's contribution to the fluency of the interaction, with a clearly significant increase:  $\mu = 3.52$ ,  $\sigma = 1.28$ ; with options  $\mu = 5.11$ ,  $\sigma = 1.42$ ;  $F = 18.6$ ,  $df = 53$ ,  $\eta^2 = 0.264$ ,  $t(26) = -4.48$ ,  $p < 0.001$ . As for the human's responsibility, i.e., how attentive to the task must be the human to correct the robot, there is a statistically significant reduction (without  $\mu = 5.30$ ,  $\sigma = 1.32$ ; with  $\mu = 4.26$ ,  $\sigma = 1.35$ ;  $F = 12.1$ ,



**Fig. 11** Assessment of the main aspects involved in the interaction. Comparison between exp. 3 in blue and exp. 4 and 5 together in red. Valuation from 1 (very low) to 7 (very high). Statistical significance marked with \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . Error bars represent std. dev.



**Fig. 12** Trust in the robot in each situation and direct comparison between third and fourth experiment. Left - Trust in the robot for the three first experiments and for exp. 4 and 5 together (both with explicit intention options). Statistical significance marked with \* $p < 0.05$ ,

\*\* $p < 0.01$ , \*\*\* $p < 0.001$ . Error bars represent std. dev. Right - Election made by the volunteers instead of valuate aspects numerically. The maximum is 27 as it is the number of volunteers. (Color figure online)

$df = 53, \eta^2 = 0.188, t(26) = 3.85, p < 0.001$ ), as well as in the importance the human gives to their own contribution to the task (without  $\mu = 5.52, \sigma = 1.34$ ; with  $\mu = 4.48, \sigma = 1.31$ ;  $F = 9.96, df = 53, \eta^2 = 0.161, t(26) = 3.31, p = 0.0027$ ). This implies that allowing the human to state their intention explicitly allows to reduce the mental burden the task places on them, at least subjectively.

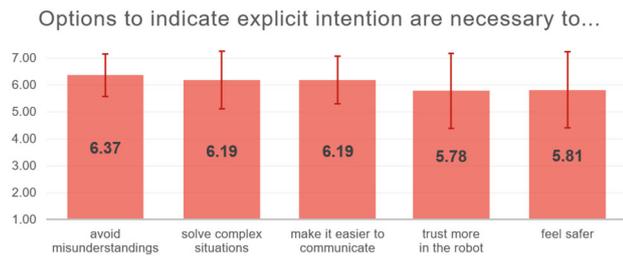
There is also a statistically significant increase in the contribution to the task that the human believes the robot makes (without  $\mu = 3.63, \sigma = 1.27$ ; with  $\mu = 4.96, \sigma = 1.43$ ;  $F = 13.1, df = 53, \eta^2 = 0.201, t(26) = -3.95, p < 0.001$ ), as well as in the belief that the robot contributes to the same extent as the human (without  $\mu = 2.70, \sigma = 1.07$ ; with  $\mu = 3.96, \sigma = 1.56$ ;  $F = 19.1, df = 53, \eta^2 = 0.271, t(26) = -4.57, p < 0.001$ ). However, the robot is still considered to contribute less than the human.

The last two parameters to be assessed are aimed at checking whether the robot is trustworthy. To do this, firstly, the volunteers are asked if they trust the robot to do the right thing when necessary, producing a statistically significant increase: without  $\mu = 3.96, \sigma = 1.53$ ; with  $\mu = 5.22, \sigma = 1.48$ ;  $F = 18.8, df = 53, \eta^2 = 0.265, t(27) = -4.49, p < 0.001$ . Second, they are asked to directly rate their degree of trust in the robot, as they did at the end of the

first and second experiments. Figure 12 Left shows the four ratings obtained.

Applying an ANOVA test we obtain that there is a statistically significant change between the four cases studied:  $F = 7.45, df = 107, \eta^2 = 0.175, p < 0.001$ . By applying a Tukey HSD (Honestly-Significant-Difference) test we can make multiple comparisons between each pair of variables. Thus, we find that there is an increase in trust in the robot when the human takes control of the task but it is not statistically representative (robot master  $\mu = 4.93, \sigma = 1.24$ ; human master  $\mu = 5.63, \sigma = 1.01$ ;  $p = 0.189$ ). In contrast, the drop in trust that occurs in the third experiment is significant indeed: human master  $\mu = 5.63, \sigma = 1.01$ ; collaborative (without options)  $\mu = 4.00, \sigma = 1.52$ ;  $p < 0.001$ . This trust is partially recovered when we give the human a way to recover part of his ability to control the task such as options to explicitly indicate his intention: collaborative (without options)  $\mu = 4.00, \sigma = 1.52$  collaborative (with options)  $\mu = 5.04, \sigma = 1.31$ ;  $p = 0.019$ .

In addition to asking the volunteers to evaluate the above parameters numerically, at the end of the fourth experiment, avoiding the possible effect that the fifth experiment could have, they were asked to choose between the third and fourth

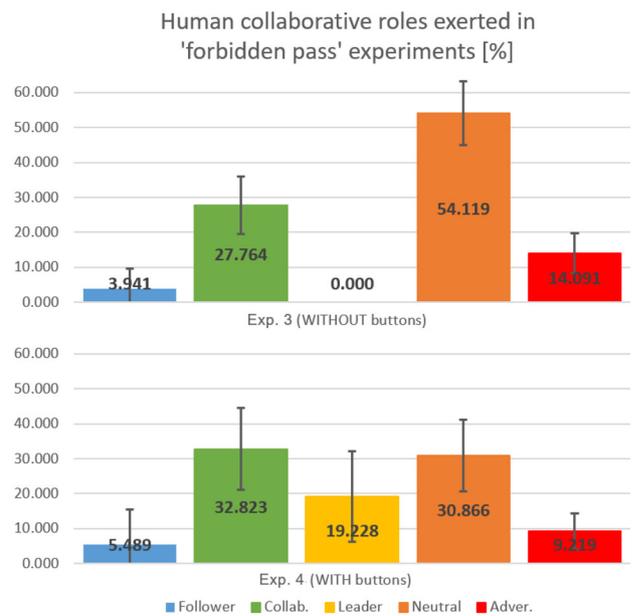


**Fig. 13** Valuation of the utility of indicating explicit intention. Valuation from 1 (totally disagree) to 7 (totally agree) with respect to different aspects. Error bars represent std. dev

experiments which of them they found safer, easier to perform, etc. The result is shown in Fig. 12 Right.

There is an unanimous response that being able to indicate your intention explicitly makes the task safer or, at least, that it does not produce any additional danger. Similarly, most volunteers feel that they find it easier and faster to execute the task when they have the options to indicate their intention and that the task runs more smoothly. On the other hand, there is disagreement as to which method makes the interaction more natural or similar to how two humans would do it. This is because both metrics depend heavily on the volunteers' ability to make the equivalence between the communication method they used (buttons) and speaking or gesturing to their partner. The statements of some of the volunteers in the post-questionnaire conversation confirm these results. "I feel safer knowing that I can always take control" says volunteer 3 regarding safety while volunteer 12 states "The robot should be able to detect that sign, a human does" regarding the naturalness of the interaction. Nevertheless, hypothesis H2 is confirmed.

As for the third hypothesis, to validate it, a final request is made to the volunteers in the last section of the questionnaire once all the experiments have been carried out. This consists of rating the usefulness of explicitly indicating their intention for various purposes. Since this last section of the questionnaire is not compared with any previous section, and in order to check if the answers are statistically significant, the result obtained for each question will be compared using ANOVAs with the scale mean, 4 (the answers range from 1 to 7), since this would be the answer they would give if they neither agreed nor disagreed with the statements in the questionnaire. The variance associated with this hypothetical response of 4 in each sentence is the same as that obtained for each sentence in the questionnaires. Fig. 13 shows that they fully agree that explicit intention is necessary both to solve complex situations ( $F = 55.3$ ,  $df = 53$ ,  $p < 0.001$ ,  $\eta^2 = 0.516$ ) and to avoid misunderstandings ( $F = 122.2$ ,  $df = 53$ ,  $p < 0.001$ ,  $\eta^2 = 0.701$ ). Likewise, they agree that this facilitates communications with the robot ( $F = 82.6$ ,  $df = 53$ ,  $p < 0.001$ ,  $\eta^2 = 0.614$ ), which allows them to



**Fig. 14** Comparison of the roles exerted by the volunteers in exp. 3 and 4. Percentage of each role detected with our system. Buttons in the handle are disabled in the third and enabled in the fourth. Error bars represent std. dev

trust it more ( $F = 22.5$ ,  $df = 53$ ,  $p < 0.001$ ,  $\eta^2 = 0.302$ ) and feel safer ( $F = 22.5$ ,  $df = 53$ ,  $p < 0.001$ ,  $\eta^2 = 0.301$ ). This reaffirms hypothesis H2 and confirms hypothesis H3.

Finally, this first round of experiments gives us a first glimpse of the role assignment allowed by our system (see Fig. 14). According to it, the human can only assume the role of leader if they explicitly indicate to the robot that they wish to take control. This, together with the other options of explicitly indicating their intention, results in the human being labeled primarily as an adversary or a neutral (uncatalogable) agent in the third experiment. In experiment 4, on the other hand, the robot knows their intention so that they can be assigned a collaborative role most of the time. It is worth noting the enormous variance in the assignment of each role, demonstrating the usefulness of the five roles, since some volunteers want to collaborate with the robot, others prefer to take the lead, and others just go along with the robot.

### 5.3 Comparison with a Classical Approach

In the previous round of experiments, it has been proven that humans agree to state their intention explicitly without undue extra effort and that they understand that this allows them to avoid misunderstandings and to resolve complex situations. However, it has not been shown that humans prefer this system to one in which only their intention is elicited through inference methods. Nor have the possible advantages of combining both types of intention been shown. At the same time, the inference of the human's implicit intention has been made

by looking only at the current force being exerted (which corresponds to the second stage of the SA) but no projection has been made into the future of this (which would correspond to the third stage of the SA).

In consequence, a second round of experiments is performed in which a force predictor is used to obtain a prediction of the force to be exerted by the human based on the environment in which they are located and the force they have exerted during the last few seconds. This is a frequent strategy in the literature: using a predictor of the human's movement to make the robot anticipate it [7, 27, 29, 88]. This predictor allows us to compare three systems: 1) a classical approach that makes use of this predictor to predict the human's intention without giving them the possibility to express it explicitly, 2) the approach of the previous experiments in which no use is made of this predictor but the human is allowed to indicate their intention, 3) a system that combines both approaches using the predictor to improve its inference but taking into account the intention explicitly given to it by the human.

In order to make this comparison, we use a new scenario (see Fig. 7 - f) specially designed so that there are much more than two different routes allowing the human to make several decisions throughout the same experiment so that they can check several times the advantages and disadvantages of each approach. In turn, we designed our own force predictor using an architecture based on a combination of Convolutional Neural Networks (CNN) [96] and Long Short-Term Memory (LSTM) [97] units. The technical details of this predictor can be found in [98].

This model uses as inputs the measurements made during the last 2 seconds of: the LiDAR of the robot, the total environmental force  $F_{E,C}^{norm}$  from (7), the linear and angular velocity of the robot and the force exerted by the human,  $f_{h,C}$  from (10). And as output, it generates a prediction of the force exerted by the human during the next second. To train it, the recorded data from the previous experiments were used. However, a sixth different scenario was used for the set of experiments, which led us to perform a preliminary round of experiments to obtain a training dataset with the aim of achieving higher accuracy compared to a predictor built only over the data from the previous five scenarios.

A total of 13 new volunteers (age:  $\mu = 31.28$ ,  $\sigma = 8.61$ ) perform 39 experiments (3 each) in which they perform the same collaborative object transportation task in the new scenario using the robot the same control structure used in the previous experiments but without allowing the human to indicate their intention explicitly in any way (equivalent to the situation shown in Exp. 3). Finally, these new samples together with the previous ones allow to obtain a predictor with an accuracy in the testset ranging from 94.4% for the force the human is about to exert in the next sampling period to 92.3% for the force to be exerted in 1 s. The dataset gener-

ated for training the force predictor used in the current study are available from the corresponding author on reasonable request.

We will use this predictor to try to anticipate whether the human wishes to go straight or turn, and if so, which way and with what intensity. This enables us to infer the position towards which the human wishes to go and use it to condition the starting point of the global planner that generates the waypoints used to generate the attractive force of the force-based model explained in Section 4.1. In this way, this system can infer the implicit intention of the human and adapt to it. As for the explicit intention, in this round of experiments it will also be obtained from the buttons on the handle of the transported bar. However, and in order to be on equal footing with the previous predictor, the options given to the human will be to indicate to the robot their intention to "take the left path", "continue through the middle" or "take the right path" (see Fig. 6c); which may allow the robot to directly discard the rest of the options when the predictor is not available or to know which way to continue after that first second if they are used together.

With these tools, a new round of experiments is designed in which 18 new volunteers (age:  $\mu = 29.44$ ,  $\sigma = 7.67$ ) perform a total of 72 experiments (4 each). Since we are dealing with a different population sample, the first experiment serves as a baseline for statistical purposes (and to train the new volunteers in the system's usage). Therefore, neither the predictor nor the buttons on the handle are used. In the second and third experiments, the same task is performed in the same scenario but adding in one of them the predictor and enabling in the other the options to indicate the intention explicitly. The order of these two experiments alternates randomly from one volunteer to the next one to avoid that this may affect their assessment. At the end, 9 volunteers performed first the experiment with the predictor and then with the buttons and the other 9 in reverse order. Finally, the fourth experiment makes use of both systems together. In order to obtain subjective measures, at the end of each of the four experiments, the volunteers completed a section of a new questionnaire.<sup>4</sup> In this case, our hypotheses to be tested are the following:

**H4** - Humans prefer to state their intention explicitly rather than the robot try to infer it.

**H5** - A system that takes into account both types of intention improves multiple parameters of effective HRI over just considering either type of intent separately.

**H6** - Humans prefer a system that takes into account both types of intentions.

Before presenting statistical results, another post-hoc statistical power test should be performed to check which is the smallest effect size that can be detected with a statistical

<sup>4</sup> 2nd round example: <https://youtu.be/cyrLoJX13mY>.

**Table 2** Objective measures associated with second round of experiments

Measurement		Value
Exp. 1 (baseline)	Duration [s]	58,23 (17,05)
	Mean force [N]	6,80 (2,46)
	Max force [N]	16,33 (7,80)
Exp. 2 or 3 (with predictor)	Duration [s]	60,83 (24,40)
	Mean force [N]	8,00 (2,67)
	Max force [N]	19,38 (9,56)
Exp. 3 or 2 (explicit intent)	Duration [s]	72,92 (19,68)
	Mean force [N]	5,15 (2,01)
	Max force [N]	13,83 (5,84)
Exp. 4 (both systems)	Duration [s]	66,86 (20,40)
	Mean force [N]	6,66 (1,92)
	Max force [N]	22,37 (7,41)

power of at least 80%. In this case we have 18 volunteers performing 4 experiments each, so if we maintain the criterion of  $p < 0.05$ , any effect size greater than  $\eta^2 = 0.145$  will have the desired statistical power.

Taking advantage of the fact that all the experiments in this round are performed in the same scenario, we can analyze objective data under equal conditions. Table 2 shows a summary of the variables considered and their mean (std. dev.) for each experiment. The large number of possible routes in this scenario causes a high variability in the duration of each experiment as the human does not always choose the shortest route. This is in addition to the natural tendency of each volunteer to perform more or less force resulting in a higher or lower speed. This variability means that the variable "Duration" does not meet the Saphiro-Wilk test, making it necessary to perform a non-parametric test. A Mann–Whitney U test is performed without obtaining significant results. The same happens with the high variability of the maximum force. Another U-test is performed without finding significant results either.

The mean force, on the other hand, does meet the normality condition. An ANOVA test is performed and a statistically significant variation is found:  $F = 3.129$ ,  $df = 71$   $p$ -value = 0.035,  $\eta^2 = 0.176$ . A Tukey HSD (Honestly-Significant-Difference) test is therefore applied to verify that there is a reduction in the mean force exerted between Exp. 2 and 3: with predictor  $\mu = 8.00$ ,  $\sigma = 2.67$ ; with buttons  $\mu = 5.15$ ,  $\sigma = 2.01$ ;  $p = 0.019$ . This reaffirms hypothesis H1 from the previous round, i.e., there is a reduction in human effort when allowed to explicitly state their intention. However, it cannot be asserted that this has an impact on faster task execution.

To test hypothesis H4, at the end of Exp. 3, the volunteers are asked through the questionnaire to choose between the system with predictor or the one that allows them to indicate

their intention explicitly in terms of parameters similar to those used in the previous round of experiments (see Fig. 12 right). They are also asked which mode of operation seems more appropriate for the task they are performing. The results are shown in Fig. 15.

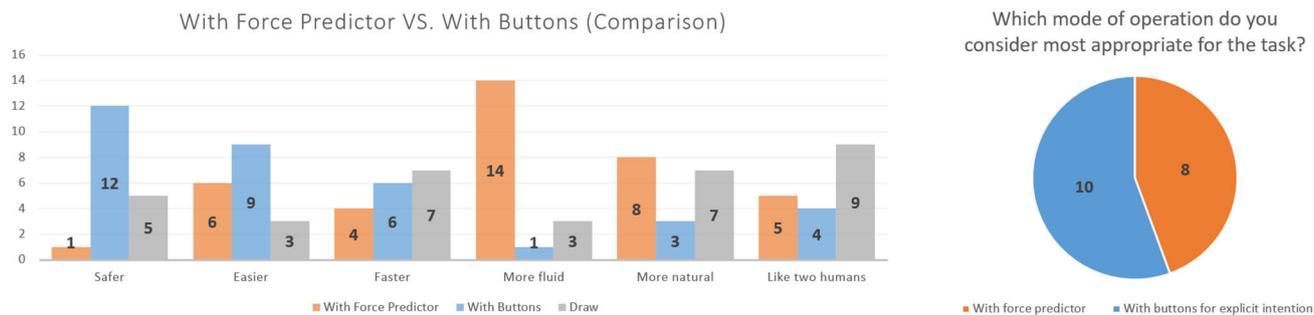
Volunteers find it safer and easier to execute the experiment in which they can express themselves explicitly. Likewise, they consider that the experiment that allows the smoothest interaction is the one in which the force predictor is used to infer their intention. On the other hand, there is no consensus as to which system allows them to execute the task faster or which system is more similar to the behavior that two humans would exhibit. Finally, there is a technical tie when it comes to choosing which system they find most appropriate, so hypothesis H4 is rejected since there is no majority (i.e. greater than two-thirds) in favor of the system that makes use of the buttons.

To test hypothesis H5, at the end of each of the four experiments in this round, volunteers are asked to fill out a section of the questionnaire (in addition to the choice part asked at the end of the third experiment and shown in Fig. 15) in which they numerically rate the same parameters that were asked in the previous round (see Fig. 11). In addition, we added two parameters that we considered could be useful, such as the quality of the solutions offered by the robot to solve the task and how comfortable the interaction with the robot was in general terms. The results obtained are shown in Fig. 16.

To obtain the statistical significances shown in Fig. 16, an ANOVA test was applied to each parameter analyzed and, if it was found that there was a significant difference among the four experiments, a Tukey HSD test was applied to determine the significance between each pair of variables. For the sake of clarity, Table 3 shows a summary of the  $p$ -values obtained for each pair of variables with a statistically significant difference.

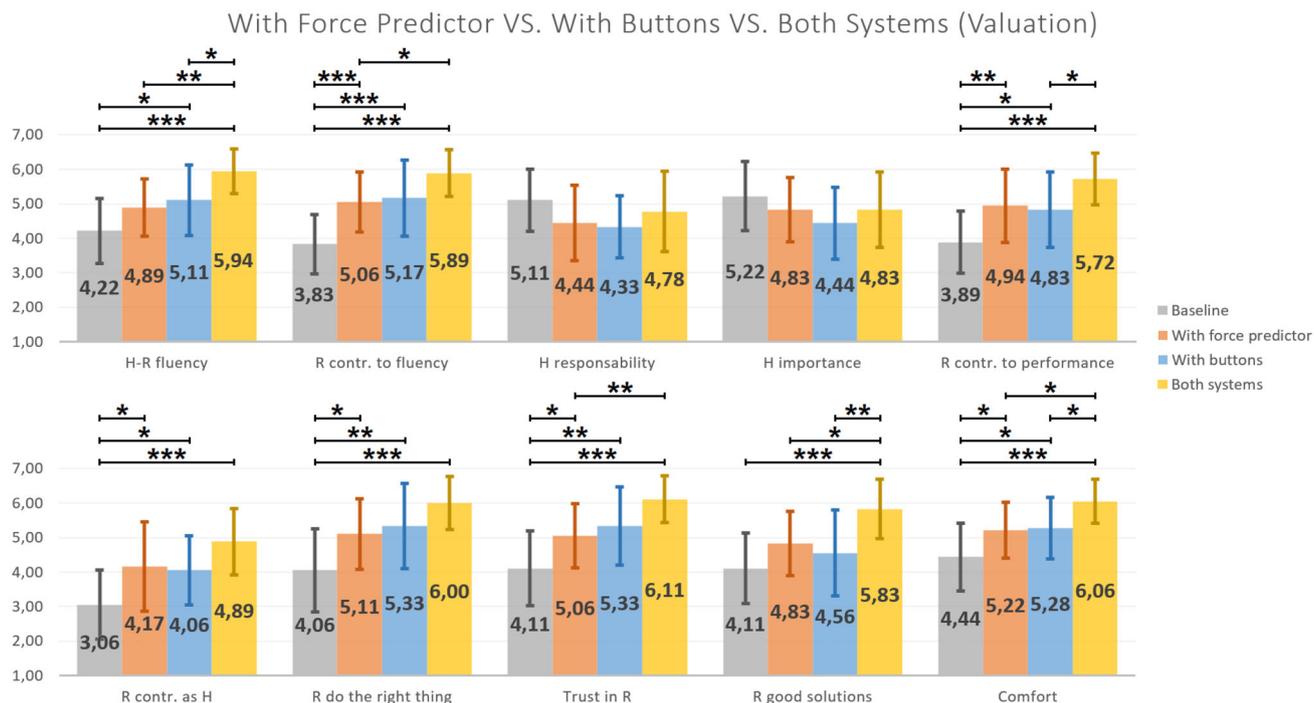
In general terms, the results presented in Table 3 show that in both the experiment using the force predictor and the experiment allowing the human to explicitly state their intention, there is a higher or lower significant improvement in practically all parameters compared to using neither of these tools, except for the human's responsibility and the importance that the human gives to their contribution to the task where no significant change is detected. This tells us that allowing the human to express their intention explicitly can achieve the same beneficial effect as using a complex predictor to try to infer their intention using significantly more computational resources. At the same time, it is also clear that the highest rated system in practically all the aspects analyzed is one that makes use of both tools together. Hypothesis H5 is thus demonstrated.

Among the results obtained, it should be noted that only the system that makes use of both types of intentions makes the human begin to consider that the robot contributes to the



**Fig. 15** Direct comparison between using force predictor and explicit intention. Left - Election made by the volunteers with respect to same aspects valuated in previous round of experiments. Right - Election

made by the volunteers with respect to which system they consider performs better at the task at hand. The maximum is 18 in both cases as it is the number of volunteers



**Fig. 16** Assessment of the main aspects involved in the interaction. Comparison among the baseline experiment (without predictor or explicit intention buttons) in gray, experiment with the force predictor in orange, experiment with the buttons in the handle in blue and

experiment with both systems together in yellow. Valuation from 1 (very low) to 7 (very high). Statistical significance marked with \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . Error bars represent std. dev. (Color figure online)

task to the same extent as the human (rating of 4.89). It is also interesting the statistically significant increase in trust in the robot that occurs when we compare the system with both tools and the one using only the predictor. This result shows that allowing the human to express themselves, even in a rudimentary way, can offer low-cost solutions to increase the degree of trust in human-robot interactions. Likewise, the complete system is able to significantly improve the human's evaluation of the solutions proposed by the robot, as well as offering the highest degree of comfort.

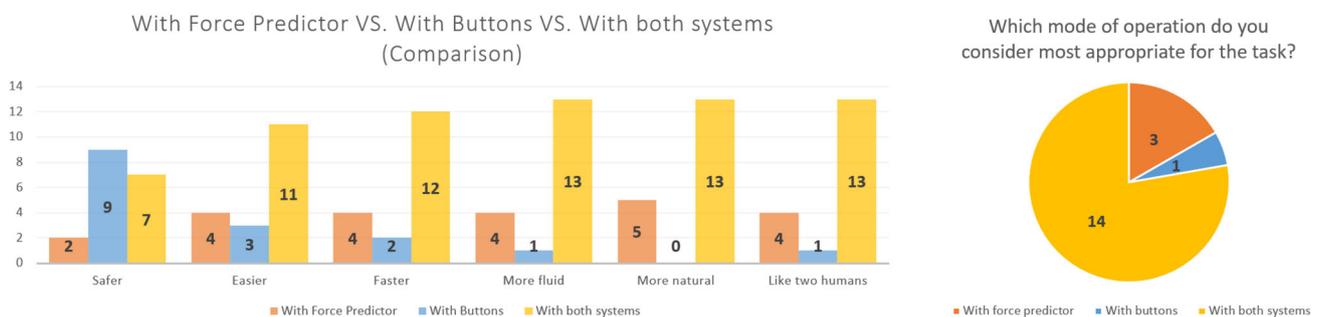
It is worth mentioning that the use of a smaller population sample in this round of experiments makes these evaluations more sensitive to some outliers making it more difficult to reach higher statistical significance as in the case of the previous round (see Fig. 11). In any case, the values obtained for the "Baseline" and "With buttons" cases are consistent with those previously obtained giving greater validity to the results of this second round of experiments.

Finally, to test hypothesis H6, we use the same method used to test hypothesis H4. At the end of the four experiments, in the last section of the questionnaire, the volunteers

**Table 3** Parameters with statistically significance in second round of experiments

Parameter	F, $\eta^2$ , <i>p</i> value	Pair analysis	
		Experiments pair	<i>p</i> value
H-R fluency	11.07, 0.328, < 0.001	Baseline - Buttons	0.016
		Baseline - Both	< 0.001
		Predictor - Both	0.003
		Buttons - Both	0.027
R contr. to fluency	9.40, 0.293, < 0.001	Baseline - Predictor	< 0.001
		Baseline - Buttons	< 0.001
		Baseline - Both	< 0.001
		Predictor - Both	0.032
H respon	0.299, 0.013, 0.826	-	-
H importance	2.46, 0.098, 0.070	-	-
R contr. to performance	6.99, 0.236, < 0.001	Baseline - Predictor	0.008
		Baseline - Buttons	0.022
		Baseline - Both	< 0.001
		Buttons - Both	0.035
R contr. as H	5.876, 0.206, 0.001	Baseline - Predictor	0.014
		Baseline - Buttons	0.033
		Baseline - Both	< 0.001
R do the right thing	9.55, 0.296, < 0.001	Baseline - Predictor	0.023
		Baseline - Buttons	0.004
		Baseline - Both	< 0.001
Trust in R	12.18, 0.350, < 0.001	Baseline - Predictor	0.025
		Baseline - Buttons	0.002
		Baseline - Both	< 0.001
		Predictor - Both	0.009
		Buttons - Both	0.002
R good solutions	9.16, 0.288, < 0.001	Baseline - Both	< 0.001
		Predictor - Both	0.023
		Buttons - Both	0.002
Comfort	10.99, 0.327, < 0.001	Baseline - Predictor	0.035
		Baseline - Buttons	0.021
		Baseline - Both	< 0.001
		Predictor - Both	0.021
		Buttons - Both	0.035

Pair analysis with Tukey HSD only performed if the analysed parameter shows a statistically significant difference (*p* < 0.05)



**Fig. 17** Direct comparison among the three systems: force predictor, explicit intention and both together. Left - Election made by the volunteers with respect to same aspects valued previously. Right - Election

made by the volunteers with respect to which system they consider performs better at the task at hand. The maximum is 18 in both cases as it is the number of volunteer

are again asked to choose, this time between the system with the predictor, the system with the buttons or the system with both tools at the same time, with respect to the parameters they had previously assessed (see Fig. 15). In addition, they are also asked again which mode of operation they find most appropriate for the task, this time with the three options mentioned above. Figure 17 shows their choices.

As it can be seen, there is a technical draw as to which system they find safer between the system that only uses the buttons to indicate explicit intent and the system that uses both tools together. This allows us to deduce that it is the ability to indicate their intention explicitly that increases the human's sense of safety, which is consistent with hypothesis H2 demonstrated in the previous round. However, and unlike the previous choice, there is no longer a division between the first two systems in terms of the rest of the parameters to be evaluated, but the volunteers mostly (with majorities of two-thirds or higher) choose the system that uses both tools as the one that allows the task to be executed more quickly or the one that makes the interaction more fluid or similar to how two humans would do it. This causes the majority choice as to which system is more appropriate for executing the task to be the complete system with a majority higher than three-fourths. Thus, hypothesis H6 holds true.

However, a non-negligible minority still chose the system with only the predictor. In the post-experiment interview, one volunteer expressed their expectation that the robot would be able to infer whatever was necessary without them having to tell it their intention. "I would like the robot to be able to do the task without me having to tell it anything," said volunteer 8. It is possible that prior expectations may keep some volunteers opting for this option. Other volunteers, on the other hand, anticipate the possible errors that this may cause, such as volunteer 10 who commented "It's good that the robot can predict my intentions but it can be wrong" which encourages them to opt for the system that makes use of both tools.

## 6 General Discussion

In this study, it has been shown that the human willingly accepts to give their intention explicitly, contrary to the general thought that we would like robots to be able to infer everything by themselves. This is made explicit in the comment made by volunteer 10 in the last round of experiments, demonstrating that, although we would like not to have to give any instructions to the robot, we are aware that it may make mistakes. This explicit intention has also been shown to improve the human's subjective assessment of the human-robot interaction by making them trust the robot more, feel safer and, most importantly, begin to consider that the robot contributes to the task in equal proportion to the human. This

is what opens the door to start considering the robot as a partner and not a mere tool. In turn, it is also what justifies that we continue working on the other aspects not analyzed, such as the effect of roles.

On the other hand, the authors acknowledge that the force-based model presented here is remarkably simple. Possibly too simple, making it not as general as it should be. For example, the estimation of the human's implicit intention is done by a simple comparison between the force the human is exerting and the force one would expect the human to exert. Similarly, obtaining the human's explicit intention is done by using three buttons which, in this case, are ergonomic and simple to use as there is one object with which the human must be in contact throughout the task. Obviously, the mechanism proposed to obtain the implicit and explicit intentions may not be valid for other types of tasks. However, as indicated in the introduction, the aim of this study and the mathematical model used is to make it as simple as possible to understand, so that the reader can easily visualize all the theoretical concepts shown in Section 3. The authors leave it as future work to use more complex and potentially more accurate inference engines, as well as other explicit communication systems, such as gesture-based or natural language processing.

Similarly, multiple concepts are presented and interrelated in Section 3. However, the experiments shown in Section 5 do not demonstrate all of them but focus on those aspects that show that the approach presented here is promising, encouraging more research related with it. The authors are aware that the influence of the assignment of collaborative roles on task performance, as well as the choice of different strategies that these roles enable, has not been analyzed in depth. We believe that these two topics deserve two specific rounds of experiments to demonstrate their promising possibilities, and are therefore beyond the scope of this first study.

About the limitations of this study, it is worth mentioning that there are also potential problems associated with the use of explicit intention communication methods. First, if the method used is not easily understood by the human, it may cause errors in its use. For this reason, we add labels next to the buttons in order to minimize errors on the part of the human. Secondly, the processing and consequent delay associated with each method of communication must be taken into account so as not to undermine the fluidity of the interaction. Finally, there is always the possibility that any of these methods may be misused or abused and, although the goodness of the human's actions is generally assumed, this possibility should not be overlooked.

Likewise, the authors recognize four weaknesses in the experimental results. First, the experimental results obtained are, to some extent, task-dependent. This implies that it is needed to implement our full cycle on a different task and check that this also improves performance in terms similar to

those obtained here, probably using different methods to infer the implicit intention and to obtain the explicit one. At the same time, this task has been executed in a laboratory environment, which does not always capture the full complexity of the same task executed in a real-world setup. Second, the set of parameters used (weights, thresholds, gains...) has been the same in all experiments. This makes them comparable to each other but also limits our knowledge of the effect these parameters have on the task. That is, it has not been tested what effect it would have if the robot exhibited a more aggressive or a more cautious behavior. Nevertheless, it is considered that this does not undermine the fundamental objectives of the present work since the expected result of changing the set of parameters would be a reduction or increase in the subjective assessment of the related aspects.

Third, the population sample size used in the second round of experiments may be small. Although the results obtained are remarkably consistent with those obtained in the first round, an effort has been made to give sufficient implementation details in Section 4.2 to increase the reproducibility of the experiments and to allow any other researcher to implement and test our model if desired. Finally, the fourth weakness of the experimental results relates to how hypotheses H4 and H6 have been confirmed or rejected. These have been addressed by directly asking the volunteers about which system they prefer with respect to various parameters or which one they consider more appropriate for the task. This type of direct question forces the user to choose and allows at a glance to know their preferences. However, it also eliminates the possibility of performing a classical statistical analysis based on ANOVA tests, for example. Because of this, their acceptance or rejection must be carefully considered. This is a limitation of the study that we believe can be solved in future studies by changing the type of question so that they also choose between one system or the other but using a graduated scale (for example between -3 and 3 with 0 representing the draw) so that further statistical analysis can be performed.

As for the effect of the robot used, it is neither fully humanoid, as would be the Talos or ARI robots from the same manufacturer, nor of industrial type. The authors consider that the numerical values obtained in the experimental part would vary slightly when using a robot that is closer to one of these two categories, tending to be lower in some evaluations, both for the base experiment and for each subsequent one, and higher in other evaluations. However, they also believe that the statistically significant effects observed when using each system would remain the same. In other words, the human should still trust a robot that makes use of a system that allows them to communicate both explicitly and implicitly more than one that allows only one type of communication.

Finally, it is worth mentioning some ethical issues related to our work. Regarding its potential impact on human

employment, the framework proposed is oriented at all times to collaborative tasks in which at least one human is always involved, offering more than a displacement, the possibility of improving human capabilities. Regarding its potential impact on human safety and autonomy, the approach of combining the implicit intention with the explicit one directly indicated by the human allows to improve human safety by being able to control the development of the task and its related decision-making in more flexible ways. Regarding data protection, an effort has been made not to use data sources that could identify the user such as video or voice. However, if they were to be used, the users' right to privacy should be taken into account and the information should be processed in a way that does not cause any potential harm.

## 7 Conclusions

In this article we have presented our Perception-Intention-Action cycle in all its depth, which we believe allows to encompass and fit together multiple works present in the literature. Using the concept of Situation Awareness, we combine the implicit intention of the human that can be inferred by the robot with the explicit intention directly delivered by the human, thus achieving a better understanding of the current situation in which the human-robot collaboration takes place, and with that, a projection of the future situation. These enable the assignment of collaborative roles to the interacting agents.

In the experimental part, it has been shown that the human agrees to give their intention explicitly without this placing an excessive extra effort on them and understands that this allows avoiding misunderstandings and resolving complex situations. It has also been shown that using a system that allows the human to express themselves explicitly can improve the human's subjective assessment of their interaction with the robot to the same extent as using complex systems that seek to infer the human's intention as accurately as possible. This is because the human prefers to maintain some sense of control over the task and because deep down they know that the robot can make mistakes. Finally, it has been found that the system that achieves the best ratings in its interaction with the human is the one that combines the inference of the human's intention with the possibility of the human directly expressing their intention.

These three results open the door to designing systems that do not seek, or rely on using, increasingly accurate inference engines, but rather systems that improve communication with the human to be able to explicitly ask for their intention or preferences in the most human possible way. This increases trust in the robot as well as the human's sense of safety.

**Acknowledgements** The authors want to express their gratitude to Ferrán Cortés, Patrick Grosch and Sergi Hernández for their technical support and to all the volunteers who made this work possible.

**Funding** Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. Work supported under the European project CANOPIES (H2020-ICT-2020-2-101016906), the MINECO/AEI ROCOTRANSP project (PID2019-106702RB-C21 MCIN/AEI/10.13039/501100011033) and the JST Moonshot R & D Grant No. JPMJMS2011-85. The first author acknowledges Spanish FPU Grant with ref. FPU19/06582.

## Declarations

**Conflict of interest** There are none potential conflicts of interest that could bias the evaluation or results of our research.

**Ethics approval** All the experiments reported in this document have been performed under the approval of the ethics committee of the Universitat Politècnica de Catalunya (UPC) in accordance with all the relevant guidelines and regulations (ID: 2021.10).

**Consent to participate** All the volunteers who participated in this study have signed an informed consent form accepting to participate in the study.

**Consent for publication** All the volunteers who participated in this study have signed an informed consent form accepting to publish the anonymously obtained data.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix A Example of the Questions Used in Questionnaires

In this section, a sample of the questionnaires used in both rounds of experiments is shown.

Generally speaking, after each experiment in each round, each volunteer fulfills a section of the questionnaire with some 7-point likert questions followed by some multiple-choice questions in which they should choose between system A or B (being sometimes the draw also an option) if pertinent (it is not pertinent to ask for an election between two systems before testing both systems). After fulfilling the section relative to the experiment just carried out, they perform the next experiment in the round and then they fulfill the next section of the questionnaire, usually with exactly the same

questions, in order to evaluate the difference. At the end of all the experiments in the round, a research assistant makes the volunteers some open questions in which they can express themselves as desired.

The 7-point likert questions are the following (questions 9 and 10 were added in the second round of experiments.):

Using the provided scale, to what extent do you agree or disagree with the following statements? (1: definitely disagree, 7: definitely agree).

1. The human–robot team worked fluently together.
2. The robot contributed positively to fluency of the interaction.
3. I had to bear the brunt to make the human–robot team better.
4. I was the most important team member on the team.
5. The robot contributed positively to the team performance.
6. The robot contributed equally to the team performance.
7. I trusted the robot to do the right thing at the right time.
8. The robot was trustworthy.
9. The robot proposed good solutions to complete the task.
10. I felt comfortable accompanying the robot.

The multiple-choice likert questions are the following:

With respect to the (system A) and (system B) experiments, which one do you consider as...?

1. Safer.
2. Easier to execute.
3. More agile.
4. More fluid.
5. More natural.
6. More like two humans.

The open-ended questions are the following:

1. What do you think of the experiment?
2. How did you feel in each experiment?
3. What would you change and why?

## References

1. Albus JS (1975) A new approach to manipulator control: the Cerebellar Model Articulation Controller (CMAC). *Trans ASME*
2. Brooks R (1986) A robust layered control system for a mobile robot. *IEEE J Robot Autom* 2(1):14–23
3. Ordóñez FJ, Roggen D (2016) Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors* 16(1):115
4. Jain A, Singh A, Koppula HS, Soh S, Saxena A (2016) Recurrent neural networks for driver activity anticipation via sensory-fusion architecture. In: 2016 IEEE international conference on robotics and automation (ICRA), pp 3118–3125. <https://doi.org/10.1109/ICRA.2016.7487478>

5. Schydlo P, Rakovic M, Jamone L, Santos-Victor J (2018) Anticipation in human–robot cooperation: a recurrent neural network approach for multiple action sequences prediction. In: 2018 IEEE international conference on robotics and automation (ICRA). IEEE, pp 5909–5914
6. Endsley MR, Garland DJ (2000) Situation awareness analysis and measurement, vol 609. CRC Press, New York. <https://doi.org/10.1201/b12461>
7. Thobbi A, Gu Y, Sheng W (2011) Using human motion estimation for human–robot cooperative manipulation. In: 2011 IEEE/RSJ international conference on intelligent robots and systems. IEEE, pp 2873–2878
8. Bussy A, Kheddar A, Crosnier A, Keith F (2012) Human-humanoid haptic joint object transportation case study. In: 2012 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, pp 3633–3638
9. Rozo L, Bruno D, Calinon S, Caldwell DG (2015) Learning optimal controllers in human–robot cooperative transportation tasks with position and force constraints. In: 2015 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, pp 1024–1030
10. Lanini J, Razavi H, Urain J, Ijspeert A (2018) Human intention detection as a multiclass classification problem: application in physical human–robot interaction while walking. *IEEE Robot Autom Lett* 3(4):4171–4178
11. Agravante DJ, Cherubini A, Bussy A, Gergondet P, Kheddar A (2014) Collaborative human–humanoid carrying using vision and haptic sensing. In: 2014 IEEE international conference on robotics and automation (ICRA). IEEE, pp 607–612
12. Tarbouriech S, Navarro B, Fraisse P, Crosnier A, Cherubini A, Sallé D (2019) Admittance control for collaborative dual-arm manipulation. In: 2019 19th international conference on advanced robotics (ICAR). IEEE, pp 198–204
13. Yu X, Li B, He W, Feng Y, Cheng L, Silvestre C (2021) Adaptive-constrained impedance control for human–robot co-transportation. *IEEE Trans Cybern* 52(12):13237–13249
14. Helbing D, Molnar P (1995) Social force model for pedestrian dynamics. *Phys Rev E* 51(5):4282
15. Domínguez-Vidal JE, Rodríguez N, Sanfeliu A (2023) Perception–intention–action cycle as a human acceptable way for improving human–robot collaborative tasks. In: Proceedings of the 2023 ACM/IEEE international conference on human–robot interaction. <https://doi.org/10.1145/3568294.3580149>. ACM/IEEE, pp 567–571
16. Cutsuridis V, Taylor JG (2013) A cognitive control architecture for the perception–action cycle in robots and agents. *Cogn Comput* 5(3):383–395
17. Albus JS et al (1993) A reference model architecture for intelligent systems design. An Introduction to Intelligent and Autonomous Control, pp 27–56
18. Nishiwaki K, Sugihara T, Kagami S, Kanehiro F, Inaba M, Inoue H (2000) Design and development of research platform for perception–action integration in humanoid robot: H6. In: 2000 IEEE/RSJ international conference on intelligent robots and systems (IROS), pp 1559–1564. <https://doi.org/10.1109/IROS.2000.895195>
19. Bekey GA (1998) On autonomous robots. *Knowl Eng Rev* 13(2):143–146
20. Rubio F, Valero F, Llopi-Albert C (2019) A review of mobile robots: concepts, methods, theoretical framework, and applications. *Int J Adv Rob Syst* 16(2):143–146
21. Fu C, Liu C, Ishi CT, Yoshikawa Y, Iio T, Ishiguro H (2021) Using an android robot to improve social connectedness by sharing recent experiences of group members in human–robot conversations. *IEEE Robot Autom Lett* 6(4):6670–6677
22. Kaneko K, Kaminaga H, Sakaguchi T, Kajita S, Morisawa M, Kumagai I, Kanehiro F (2019) Humanoid robot hrp-5p: an electrically actuated humanoid robot with high-power and wide-range joints. *IEEE Robot Autom Lett* 4(2):1431–1438
23. Gonzalez-Aguirre JA, Osorio-Oliveros R, Rodríguez-Hernández KL, Lizárraga-Iturralde J, Morales Menendez R, Ramírez-Mendoza RA, Ramírez-Moreno MA, Lozoya-Santos JDJ (2021) Service robots: trends and technology. *Appl Sci* 11(22):10702
24. Lee B-J et al (2018) Perception–action–learning system for mobile social-service robots using deep learning. In: Proceedings of the AAAI conference on artificial intelligence, vol 32
25. Luo RC, Mai L (2019) Human intention inference and on-line human hand motion prediction for human–robot collaboration. In: 2019 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, pp 5958–5964
26. Jain S, Argall B (2018) Recursive Bayesian human intent recognition in shared-control robotics. In: 2018 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, pp 3905–3912
27. Huang C-M, Mutlu B (2016) Anticipatory robot control for efficient human–robot collaboration. In: 2016 11th ACM/IEEE international conference on human–robot interaction (HRI), pp 83–90. <https://doi.org/10.1109/HRI.2016.7451737>
28. Maroger I, Ramuzat N, Stasse O, Watier B (2021) Human trajectory prediction model and its coupling with a walking pattern generator of a humanoid robot. *IEEE Robot Autom Lett* 6(4):6361–6369
29. Alevizos KI, Bechlioulis CP, Kyriakopoulos KJ (2020) Physical human–robot cooperation based on robust motion intention estimation. *Robotica* 38(10):1842–1866
30. Evrard P, Gribovskaya E, Calinon S, Billard A, Kheddar A (2009) Teaching physical collaborative tasks: object-lifting case study with a humanoid. In: 2009 9th IEEE-RAS international conference on humanoid robots. IEEE, pp 399–404
31. Fridovich-Keil D, Bajcsy A, Fisac JF, Herbert SL, Wang S, Dragan AD, Tomlin CJ (2020) Confidence-aware motion prediction for real-time collision avoidance I. *Int J Robot Res* 39(2–3):250–265
32. Gilbert M (2009) Shared intention and personal intentions. *Philos Stud* 144:167–187
33. Bratman ME (1993) Shared intention. *Ethics* 104(1):97–113
34. Dominey PF, Wameken F (2011) The basis of shared intentions in human and robot cognition. *New Ideas Psychol* 29(3):260–274
35. Jain S, Argall B (2019) Probabilistic human intent recognition for shared autonomy in assistive robotics. *ACM Trans Hum Robot Interact* 9(1):1–23
36. Gardner R, Mancero Castillo CS, Wilson S, Farina D, Burdet E, Khoo BC, Atashzar SF, Vaidyanathan R (2020) A multimodal intention detection sensor suite for shared autonomy of upper-limb robotic prostheses. *Sensors* 20(21):6097
37. Zhuang KZ, Sommer N, Mendez V, Aryan S, Formento E, D’Anna E, Artoni F, Petrini F, Granata G, Cannaviello G et al (2019) Shared human–robot proportional control of a dexterous myoelectric prosthesis. *Nat Mach Intell* 1(9):400–411
38. Losey DP, McDonald CG, Battaglia E, O’Malley MK (2018) A review of intent detection, arbitration, and communication aspects of shared control for physical human-robot interaction. *Appl Mech Rev* 70:1. <https://doi.org/10.1115/1.4039145>
39. Vernon D, Thill S, Ziemke T (2016) The role of intention in cognitive robotics. Toward robotic socially believable behaving systems-volume i: modeling emotions, pp 15–27
40. Lyons JB, Havig PR (2014) Transparency in a human–machine context: approaches for fostering shared awareness/intent. In: Virtual, augmented and mixed reality. designing and developing virtual and augmented environments: 6th international conference, VAMR 2014, held as part of HCI international 2014, Heraklion, Crete, Greece, June 22–27, 2014, Proceedings, Part I 6. Springer, pp 181–190

41. Mullen JF, Mosier J, Chakrabarti S, Chen A, White T, Losey DP (2021) Communicating inferred goals with passive augmented reality and active haptic feedback. *IEEE Robot Autom Lett* 6(4):8522–8529
42. Che Y, Okamura AM, Sadigh D (2020) Efficient and trustworthy social navigation via explicit and implicit robot–human communication. *IEEE Trans Robot* 36(3):692–707
43. Gildert N (2022) Combining implicit and explicit communication in object manipulation tasks between two robots. Ph.D. thesis, University of York
44. Dalmasso M, Garrell A, Domínguez-Vidal JE, Jiménez P, Sanfeliu A (2021) Human–robot collaborative multi-agent path planning using Monte Carlo tree search and social reward sources. In: *IEEE International conference on robotics and automation (ICRA)*
45. Domínguez-Vidal JE, Torres-Rodríguez IJ, Garrell A, Sanfeliu A (2021) User-friendly smartphone interface to share knowledge in human–robot collaborative search tasks. In: *30th IEEE international conference on robot and human interactive communication (RO-MAN)*, pp 913–918. <https://doi.org/10.1109/RO-MAN50785.2021.9515379>
46. Gildert N, Millard AG, Pomfret A, Timmis J (2018) The need for combining implicit and explicit communication in cooperative robotic systems. *Front Robot AI* 5:65
47. Dar S, Bernardet U (2020) When agents become partners: a review of the role the implicit plays in the interaction with artificial social agents. *Multimodal Technol Interact* 4(4):81
48. Endsley MR (2000) Direct measurement of situation awareness: validity and use of SAGAT. *Situat Aware Anal Measur* 10:147–173
49. Scholtz J (2002) Evaluation methods for human–system performance of intelligent systems. Technical report, National Inst of Standards and Technology Gaithersburg MD Manufacturing Engineering Lab
50. Riley JM, Endsley MR (2004) The hunt for situation awareness: human–robot interaction in search and rescue. In: *Proceedings of the human factors and ergonomics society annual meeting*, vol 48, pp 693–697
51. Riley JM, Strater LD, Chappell SL, Connors ES, Endsley MR (2010) Situation awareness in human–robot interaction: challenges and user interface requirements. *Human–robot interactions in future military operations*, pp 171–192
52. Opiyo S, Zhou J, Mwangi E, Kai W, Sunusi I (2021) A review on teleoperation of mobile ground robots: architecture and situation awareness. *Int J Control Autom Syst* 19(3):1384–1407
53. Yanco HA, Drury J (2004) "Where am I?" Acquiring situation awareness using a remote robot platform. In: *2004 IEEE international conference on systems, man and cybernetics (IEEE Cat. No. 04CH37583)*, vol 3. IEEE, pp 2835–2840
54. Scholtz JC, Antonishek B, Young JD (2005) Implementation of a situation awareness assessment tool for evaluation of human–robot interfaces. *IEEE Trans Syst Man Cybern Part A Syst Hum* 35(4):450–459
55. Roldán JJ, Peña-Tapia E, Martín-Barrio A, Olivares-Méndez MA, Del Cerro J, Barrientos A (2017) Multi-robot interfaces and operator situational awareness: study of the impact of immersion and prediction. *Sensors* 17(8):1720
56. Dini A, Murko C, Yahyanejad S, Augsdörfer U, Hofbauer M, Paletta L (2017) Measurement and prediction of situation awareness in human–robot interaction based on a framework of probabilistic attention. In: *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, pp 4354–4361
57. Simaan N, Taylor RH, Choset H (2015) Intelligent surgical robots with situational awareness. *Mech Eng* 137(09):3–6
58. Ratwani RM, McCurry JM, Trafton JG (2010) Single operator, multiple robots: an eye movement based theoretic model of operator situation awareness. In: *2010 5th ACM/IEEE international conference on human–robot interaction (HRI)*. IEEE, pp 235–242
59. Riley JM, Endsley MR (2005) Situation awareness in HRI with collaborating remotely piloted vehicles. In: *Proceedings of the human factors and ergonomics society annual meeting*, vol 49. SAGE Publications, Los Angeles, CA, pp 407–411
60. Scholtz J, Antonishek B, Young J (2004) Evaluation of a human–robot interface: Development of a situational awareness methodology. In: *Proceedings of the 37th annual Hawaii international conference on system sciences*, 2004. IEEE, p 9
61. Ghezala MWB, Bouzeghoub A, Leroux C (2014) Rsaw: a situation awareness system for autonomous robots. In: *2014 13th international conference on control automation robotics & vision (ICARCV)*. IEEE, pp 450–455
62. Ginesi M, Meli D, Roberti A, Sansonetto N, Fiorini P (2020) Autonomous task planning and situation awareness in robotic surgery. In: *2020 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, pp 3144–3150
63. Mörtl A, Lawitzky M, Kucukyilmaz A, Sezgin M, Basdogan C, Hirche S (2012) The role of roles: physical cooperation between humans and robots. *Int J Robot Res* 31(13):1656–1674. <https://doi.org/10.1177/0278364912455366>
64. Philips J, Millán JdR, Vanacker G, Lew E, Galán F, Ferrez PW, Van Brussel H, Nuttin M (2007) Adaptive shared control of a brain-actuated simulated wheelchair. In: *2007 IEEE 10th international conference on rehabilitation robotics*. IEEE, pp 408–414
65. Dragan AD, Srinivasa SS (2013) A policy-blending formalism for shared control. *Int J Robot Res* 32(7):790–805
66. Marcano M, Díaz S, Pérez J, Irigoyen E (2020) A review of shared control for automated vehicles: theory and applications. *IEEE Trans Hum Mach Syst* 50(6):475–491
67. Ferrer G, Sanfeliu A (2014) Proactive kinodynamic planning using the extended social force model and human motion prediction in urban environments. In: *2014 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, pp 1730–1735
68. Farina F, Fontanelli D, Garulli A, Giannitrapani A, Prattichizzo D (2017) Walking ahead: the headed social force model. *PLoS ONE* 12(1):0169734
69. Yang C-T, Zhang T, Chen L-P, Fu L-C (2019) Socially-aware navigation of omnidirectional mobile robot with extended social force model in multi-human environment. In: *2019 IEEE international conference on systems, man and cybernetics (SMC)*. IEEE, pp 1963–1968
70. Garrell A, Garza-Elizondo L, Villamizar M, Herrero F, Sanfeliu A (2017) Aerial social force model: a new framework to accompany people using autonomous flying robots. In: *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, pp 7011–7017
71. Gil O, Garrell A, Sanfeliu A (2021) Social robot navigation tasks: combining machine learning techniques and social force model. *Sensors*. <https://doi.org/10.3390/s21217087>
72. Liu S, Hou Z, Tian T, Deng Z, Li Z (2019) A novel dual successive projection-based model-free adaptive control method and application to an autonomous car. *IEEE Trans Neural Netw Learn Syst* 30(11):3444–3457
73. Xu D, Zhang W, Shi P, Jiang B (2019) Model-free cooperative adaptive sliding-mode-constrained-control for multiple linear induction traction systems. *IEEE Trans Cybern* 50(9):4076–4086
74. Tutsoy O, Barkana DE, Balıkcı K (2021) A novel exploration–exploitation-based adaptive law for intelligent model-free control approaches. *IEEE Trans Cybern* 53(1):329–337
75. Raggioli L, D'Asaro FA, Rossi S (2023) Deep reinforcement learning for robotic approaching behavior influenced by user activity and disengagement. *Int J Soc Robot*. <https://doi.org/10.1007/s12369-023-01044-7>

76. Zehfroosh A, Tanner HG (2022) A hybrid PAC reinforcement learning algorithm for human–robot interaction. *Front Robot AI* 9:797213
77. Pong V, Gu S, Dalal M, Levine S (2018) Temporal difference models: model-free deep rl for model-based control. arXiv preprint [arXiv:1802.09081](https://arxiv.org/abs/1802.09081)
78. Gentner D, Stevens AL (2014) *Mental models*. Psychology Press, London
79. Yousif K, Bab-Hadiashar A, Hoseinnezhad R (2015) An overview to visual odometry and visual SLAM: applications to mobile robotics. *Intell Ind Syst* 1(4):289–311
80. Lajoie P-Y, Ramtoula B, Chang Y, Carlone L, Beltrame G (2020) Door-slam: distributed, online, and outlier resilient slam for robotic teams. *IEEE Robot Autom Lett* 5(2):1656–1663
81. Bai Y, Zhang B, Xu N, Zhou J, Shi J, Diao Z (2023) Vision-based navigation and guidance for agricultural autonomous vehicles and robots: a review. *Comput Electron Agric* 205:107584
82. Villamizar M, Andrade-Cetto J, Sanfeliu A, Moreno-Noguer F (2017) Boosted random ferns for object detection. *IEEE Trans Pattern Anal Mach Intell* 40(2):272–288
83. Vaquero V, Sanfeliu A, Moreno-Noguer F (2018) Deep lidar cnn to understand the dynamics of moving vehicles. In: 2018 IEEE international conference on robotics and automation (ICRA). IEEE, pp 4504–4509
84. Viswakumar A, Rajagopalan V, Ray T, Parimi C (2019) Human gait analysis using openpose. In: 2019 fifth international conference on image information processing (ICIIP). IEEE, pp 310–314
85. Kam M, Zhu X, Kalata P (1997) Sensor fusion for mobile robot navigation. *Proc IEEE* 85(1):108–119
86. Li C, Wang S, Zhuang Y, Yan F (2019) Deep sensor fusion between 2d laser scanner and IMU for mobile robot localization. *IEEE Sens J* 21(6):8501–8509
87. Luo R, Berenson D (2015) A framework for unsupervised online human reaching motion recognition and early prediction. In: 2015 IEEE/RSJ international conference on intelligent robots and systems (IROS), pp 2426–2433. <https://doi.org/10.1109/IROS.2015.7353706>
88. Laplaza J, Pumarola A, Moreno-Noguer F, Sanfeliu A (2021) Attention deep learning based model for predicting the 3D human body pose using the robot human handover phases. In: 30th IEEE international conference on robot and human interactive communication (RO-MAN), pp. 161–166. <https://doi.org/10.1109/RO-MAN50785.2021.9515402>
89. Jarrassé N, Charalambous T, Burdet E (2012) A framework to describe, analyze and generate interactive motor behaviors. *PLoS ONE* 7(11):1–13. <https://doi.org/10.1371/journal.pone.0049945>
90. Kucukyilmaz A, Sezgin TM, Basdogan C (2012) Intention recognition for dynamic role exchange in haptic collaboration. *IEEE Trans Haptics* 6(1):58–68
91. Lippi M, Marino A (2021) A mixed-integer linear programming formulation for human multi-robot task allocation. In: 2021 30th IEEE international conference on robot & human interactive communication (RO-MAN). IEEE, pp 1017–1023
92. Lippi M, Di Lillo P, Marino A (2023) A task allocation framework for human multi-robot collaborative settings. In: 2023 IEEE international conference on robotics and automation (ICRA). IEEE, pp 7614–7620
93. Rusu RB, Cousins S (2011) 3D is here: point cloud library (PCL). In: IEEE international conference on robotics and automation (ICRA). IEEE, pp 1–4
94. Laplaza J, Rodríguez N, Domínguez-Vidal JE, Herrero F, Hernández S, López A, Sanfeliu A, Garrell A (2022) IVO Robot: a new social robot for human–robot collaboration. In: Proceedings of the 2022 ACM/IEEE international conference on human–robot interaction. IEEE, pp 860–864
95. Zanlungo F, Ikeda T, Kanda T (2011) Social force model with explicit collision prediction. *EPL (Europhys Lett)* 93(6):68005. <https://doi.org/10.1209/0295-5075/93/68005>
96. Dhillon A, Verma GK (2020) Convolutional neural network: a review of models, methodologies and applications to object detection. *Progr Artif Intell* 9(2):85–112
97. Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780
98. Domínguez-Vidal JE, Sanfeliu A (2023) Improving human–robot interaction effectiveness in human–robot collaborative object transportation using force prediction. In: 2023 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.