



HAL
open science

Evaluating the Pertinence of Robot Decisions in a Human-Robot Joint Action Context: The PeRDITA Questionnaire

Sandra Devin, Camille Vrignaud, Kathleen Belhassein, Aurélie Clodic, Ophélie Carreras, Rachid Alami

► **To cite this version:**

Sandra Devin, Camille Vrignaud, Kathleen Belhassein, Aurélie Clodic, Ophélie Carreras, et al.. Evaluating the Pertinence of Robot Decisions in a Human-Robot Joint Action Context: The PeRDITA Questionnaire. 27th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2018), Aug 2018, Nanjing, China. 10.1109/ROMAN.2018.8525785 . hal-01906101

HAL Id: hal-01906101

<https://laas.hal.science/hal-01906101v1>

Submitted on 26 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Evaluating the Pertinence of Robot Decisions in a Human-Robot Joint Action Context: The PeRDITA Questionnaire

Sandra Devin¹, Camille Vrignaud¹, Kathleen Belhassein^{1,2}, Aurélie Clodic¹, Ophélie Carreras², Rachid Alami¹

Abstract—The domain of human-robot Joint Action is a growing field where roboticists, psychologists and philosophers start to collaborate in order to devise robot abilities that are as efficient and convenient for the human partner as possible. Besides studying Joint Action and developing algorithms and schemes to control the robot and manage the interaction, one of the current challenges is to come up with a method to properly evaluate the progresses made by the community. Several questionnaires have already been proposed to the community that deal with the evaluation of human-robot interaction. However, these studies mainly concern either specific basic behaviors during Joint Action or human-robot interactions without effective physical Joint Action. When it comes to high level decisions during physical human-robot Joint Action, there are fewer contributions to the topic, and also, the methods to evaluate them are even rarer. The aim of this paper is to propose a reusable questionnaire PeRDITA (Pertinence of Robot Decisions In jointT Action) allowing us to evaluate the pertinence of high level decision abilities of a robot during physical Joint Action with a human.

I. INTRODUCTION

In industrial as well as in human assistance contexts, the robot of the future needs to be able to perform Joint Action with humans. Joint Action is defined as "any form of social interaction whereby two or more individuals coordinate their actions in space and time to bring about a change in the environment" [1]. Our study concerns the more precise framework of two "partners" face to face, a robot and a human, sharing the common goal of a collaborative task achievement.

This domain of human-robot Joint Action is a growing field where roboticists, psychologists and philosophers start to collaborate in order to devise robot abilities that are as efficient and convenient for the human partner as possible. Besides developing algorithms and schemes to control the robot and manage its interactions, one of the current challenges is to come up with a method to properly evaluate the progresses made by the community. Since measurement of physiological data on the user could be hard to implement, a usual way to do it is through user studies where naive users answer a questionnaire about their experience of the interaction. Several questionnaires have already been proposed to the community [2], [3] and other studies used "homemade"

questionnaires [4], [5]. However, our research interest is based on physical interactions that are involved in human-robot collaborative task achievement, where the human and the robot share the task and the space. The questionnaires mentioned above mainly concern rather specific and basic behaviors or tasks without concrete interaction. When it comes to the evaluation of high level decision making processes concerning a concrete physical human-robot interaction, the methods to evaluate these contributions are even rarer.

Concerning this decisional level, the ideal teammate/assistant robot should not only make, whenever necessary, the pertinent decisions with minimum intrusion and maximum latitude offered to its human partner but should also constantly exhibit that it has a good estimate of its human partner mental state with respect to the shared task.

We are contributing to the field, and we develop incrementally architectures [6] and schemes [7], [8] for human-robot collaborative task achievement, but we did not find a questionnaire to evaluate the robot high level decision making processes in this context.

Our goal here is to propose and incrementally refine a questionnaire PeRDITA (Pertinence of Robot Decisions In jointT Action) that is not specific to a robot or a task. A pilot study has been done using a preliminary version of this questionnaire, where subjects had to interact with a PR2 robot to achieve a joint task.

In the sequel, we will first present previous work concerning the user experience and other questionnaires used to evaluate human-robot interaction in Sec. II. Then we will present the questionnaire, how we constructed it in Sec. III and how we evaluated it through a user study presented in Sec. IV. The results are presented in Sec. V. Finally, in Sec. VI, we will discuss the re-usability of the questionnaire and the difficulties to evaluate human-robot Joint Action.

II. BACKGROUND

A. Existing questionnaires

Several questionnaires have already been developed to evaluate human-robot interaction. [9] allows us to evaluate several aspects such as the "trust" in the robot or the "fluency" of the interaction. It has been used in several studies such as [10] or [11]. The Godspeed questionnaire series [2] enable us to measure the perception of the robot by the human with questions relative to anthropomorphism or perceived intelligence. However, these questionnaires are focused on the evaluation of the perception of the robot and

*This work was supported by ANR project JointAction4HRI (ANR-16-CE33-0017)

¹LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France
name@laas.fr

²CLLE, Université de Toulouse, CNRS, Toulouse, France
name@univ-tlse2.fr

does not deal with the evaluation of the interaction and the usability of the system. The SUS (System Usability Scale) questionnaire [12] measures the interaction of a user with an electronic system with 10 claims that subjects need to evaluate using Likert scale. On the contrary of the Godspeed series, the SUS questionnaire evaluates the usability of the system but lacks of measure concerning the perception of the robot or the interaction. [3] presents a toolkit to measure acceptance for assistive social robots. This toolkit is based on the UTAUT (Unified Theory of Acceptance and Use of Technology) questionnaire [13]. It has been well conceived in order to evaluate the perception and usability of the robot and more particularly for social robots. However, the questionnaire is more oriented toward the perception of the robot than the interaction and the collaboration. The Robotic Social Attributes Scale [14], which was inspired by the Godspeed questionnaire and the literature from social psychology, measures the judgment and the perception of people in relation to the social attributes of the robot, like warmth or competence. Finally, several studies as [4], [5] use "homemade" questionnaire conceived for their experiment and not always easily reusable.

B. Need for a new questionnaire

We saw in the previous subsection that several questionnaires already exist to evaluate human-robot interaction. However, they mainly focus either on some specific basic behaviors or on evaluating human-robot interaction without concrete physical interaction. Even if these questionnaires are interesting in their respective fields of application, when it comes to evaluate high level decisions of a robot, few works have been done on the subject.

The decisions that we consider here correspond to physical (e.g. pick&place) and verbal actions that are involved when a human and a robot have to satisfy a joint goal: what action to perform, who will do and when. Indeed, when the task is a little complex, when various ways exist to achieve a same goal or when the spatial resource itself is shared by the human and the robot, it is important for the human to have a robot partner:

- that tracks and is permanently aware of the current state of the task
- that can comply with the human decisions
- that makes explicit its internal state
- that does all this with minimal intrusive behavior

The objective of researchers who contribute to the development of the robot high level decision abilities for Human-robot joint action is to come up with a robot that is able make the right decisions at the right time.

Our aim here is to provide a questionnaire partially inspired by UX models, which allows us to evaluate the progresses made by the community concerning the pertinence of high level decision abilities of the robot during human-robot Joint Action.

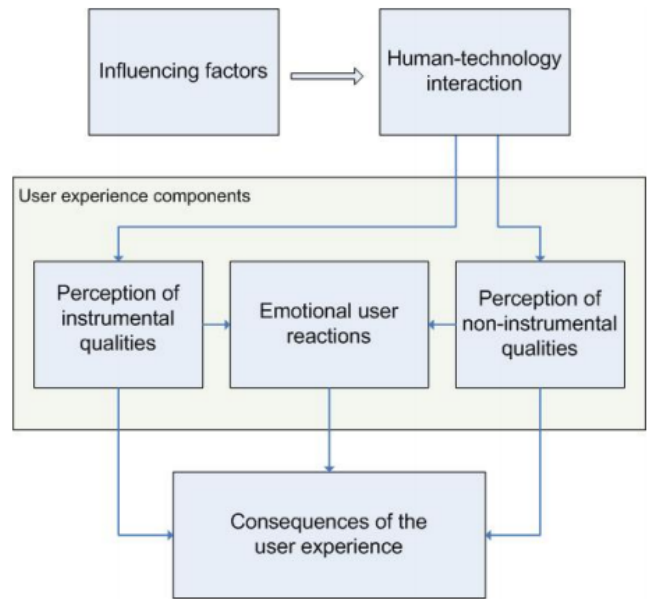


Fig. 1. UX model by [16]. The user experience is based on the human-technology interaction and is composed of three parts: the perception of instrumental qualities, the perception of non-instrumental qualities and the emotional user reaction.

C. The UX Model

The acceptability of complex technological systems as computer or robots is studied by researchers in social sciences. To do so they define what they call the User eXperience [15] as:

”a consequence of a user’s internal state (predispositions, expectations, needs, motivation, mood, etc.), the characteristics of the designed system (e.g. complexity, purpose, usability, functionality, etc.) and the context (or the environment) within which the interaction occurs (e.g. organisational/social setting, meaningfulness of the activity, voluntariness of use, etc.)”

This definition has been designed as a model by [16] as presented in Fig. 1. This model is based on the human-technology interaction and is composed of three parts: the perception of instrumental qualities, the emotional user reaction and the perception of non-instrumental qualities.

a) *Perception of instrumental qualities*: The instrumental qualities of a technology, or also called pragmatic attributes, are strongly linked to the acceptability of the technology defined in [17] as:

”the demonstrable willingness within a user group to employ information technology for the task it is designed to support.”

b) *Perception of non-instrumental qualities*: In the definition of [18], the non-instrumental qualities of a technology, or also called hedonic attributes, depend of the user and refer to the pleasure obtained by the use of the technology. It includes several notions such as the stimulation procured during the interaction, identification mechanisms and representations. This aspect is evaluated through the perception of

the technology by the user. In the human-robot interaction context, the criteria to take into account are the aesthetics of the robot, symbolic aspects and motivational aspects during the interaction.

c) Emotional user reaction: The emotions of the user after the interaction with the technology will impact the final use of the technology. A positive emotion will support a future use while a negative emotion can lead to a reject of the technology.

III. CREATION OF THE QUESTIONNAIRE

A. The dimensions

In order to evaluate the user experience concerning the robot and the interaction, we have chosen to build a questionnaire where subjects have to place themselves in a self-assessment scale. This kind of questionnaire is often used in HRI because it provides quantitative measures on which it is possible to make statistical analysis. We organized the questionnaire on several dimensions, each one measuring a specific aspect of the interaction.

- *Evaluation of perception of the instrumental qualities:* The **Interaction** dimension, based on the French version of the AttrakDiff questionnaire [19] allows us to evaluate the behavioral intention of use.
- *Evaluation of the perception of non-instrumental qualities:* we based this part on the Godspeed questionnaire series[2]. It allows us to evaluate how the human perceived the robot in general. The associated dimension of the questionnaire will be called the **Robot perception** dimension afterward.

In addition to these dimensions, we have added three other dimensions that are more specific to the context of high level decision and physical human-robot Joint Action:

- **Verbal:** this dimension allows us to evaluate how the human perceived the verbal interaction with the robot (did the robot verbalized the good information at the right time).
- **Acting:** this dimension enables us to evaluate how the human perceived the decisions of the robot concerning its actions (did the robot choose to perform the right actions at the right time).
- **Collaboration:** this dimension measures how the subject perceived the collaboration with the robot in terms of acceptability, usability, or security.

As our research interests are more about the pertinence of high level decision abilities of the robot in a context of Joint Action, we focused on the interaction parameters and not on the emotional user reaction. However, in order to better fit with the UX models, it might be interesting to add in this questionnaire this specific dimension, which is possible by using existing scales already validated such as the Self-Assessment Manikin [20] or the AffectButton [21].

B. The questions

Several antonym items are used by dimension (between 3 and 8). Subjects have to answer a question by placing

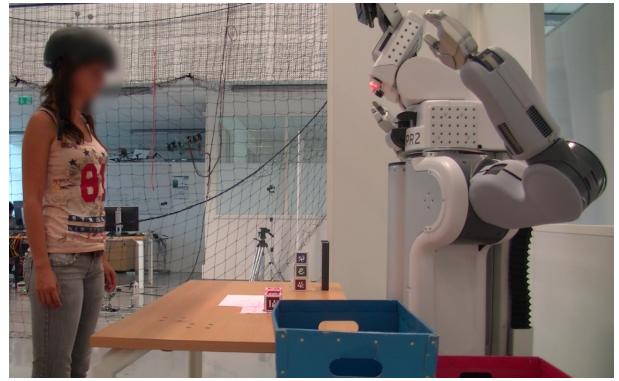


Fig. 2. A joint activity between a PR2 robot and a subject. They have to achieve together an "Inventory" task where they have to scan several colored cubes and store them into a box of the same color. Only the human can perform the scanning. In the situation depicted by the photo, the robot is scanning a cube before storing it.

themselves in a scale of 100 between these antonym items (semantic differential). For a better understanding, the English version of the questions and items can be found in Tab. I.

IV. THE PILOT STUDY

The questionnaire presented in this paper has been tested in the context of a pilot study where subjects had to perform a joint task with a PR2 robot.

A. Aim of the study

The aim of the study used to test the questionnaire was to compare two versions of a system allowing to execute Shared Plans in the context of human-robot collaboration. The first version was the initial system presented in [22] and the second version was the same system with two major improvements concerning Shared Plan elaboration and execution. The first improvement consists in taking into account an estimation (by the robot) of the mental state of its human partner during Shared Plan execution in order to better communicate concerning potential divergent beliefs [7]. The second improvement was to consider more flexible Shared Plans where the robot does not decide everything in advance but defers some decisions concerning actions allocation and instantiation to the last moment in order to provide more latitude to the human [8]. The pilot study also had the purpose to compare two different modes of last moment decision implemented in the new system. The modes and the two systems will be presented in more detail below.

B. The task

During the study, the subject and the robot have achieved together an "Inventory" task where they had to scan several colored cubes and stored them into a box of the same color. At the beginning of the interaction, both agents had a stack of colored cubes they could access (and only them could access). The cubes were blue, green or red. The stack of the human was located in another room, in a way that, to get a new cube, the human had to leave the sight of view of

TABLE I
 QUESTIONS OF EACH DIMENSION OF THE QUESTIONNAIRE. SUBJECTS HAVE TO PLACE THEMSELVES BETWEEN THE TWO
 ANTONYM ITEMS IN A SCALE OF 100.

Dimension	Question	Item
Collaboration	In your opinion, the collaboration with the robot to perform the task was:	Restrictive/Adaptive Useless/Useful Unsettling/Satisfactory Annoying/Acceptable Insecure/Secure
Interaction	In your opinion, generally, the interaction was:	Negative/Positive Complicated/Simple Not practical/Practical Unpredictable/Predictable Ambiguous/Clear
Robot perception	In your opinion, the robot is rather:	Machinelike/Humanlike Artificial/Living Inert/Animated Apathetic/Responsive Unpleasant/Pleasant Disagreeable/Agreeable Stupid/Intelligent Incompetent/Competent
Verbal	In your opinion, the robot verbal interventions were:	Incomprehensible/Clear Insufficient/Sufficient Superfluous/Pertinent
Acting	In your opinion, the robot actions were:	Inappropriate/Appropriate Useless/Useful Unpredictable/Predictable

the robot (see Fig. 4). For a cube to be scanned, the agents needed to put it on one out of the two possible areas on the table in front of the robot (see Fig. 4). Once a cube was placed on a scanning area, the robot could scan it by orienting its head and turning on a red light in the direction of the object (see Fig. 2). If the robot scanned an object while the human was not looking at the scene (e.g. when he was in another room to pick a new cube), the human was not aware that the object had been scanned unless the robot told him. Once the cube was scanned, it could be stored in a box of the same color (e.g. the blue cubes in a blue box). The robot had access to a blue box, the human to a green box, and both had access to a red box. Consequently, only the robot could store the blue cubes, only the human could store the green cubes and both could store the red cubes. The two boxes of the human were located in another room (see Fig. 4).

This task required the execution of a Shared Plan between the human and the robot and contained three main interesting points concerning the execution of this Shared Plan. First, since the cubes and the boxes of the human were located in another room, the action of the robot to scan a cube could be missed by the human and it could result a lack of knowledge that would need to be solved. Second, the presence of the red cubes that could be stored in a box by both agents required a decision: who would store a given red cube? Finally, the presence of two scan areas which were identical and could be used in the same way required the agent to take decision concerning which area to use. However, after few pretests, we noticed several possible problems that we fixed by introducing small adaptations to the task:

- *Introduction of a red video tape box:* In certain cases, the configuration coupled to the decision of the subjects led to not having any decision in the task concerning the red cubes. Indeed, there were cases where, each time there was a red cube to store, one of the two agents were busy (either the human was in another room to pick or store an object or the robot was performing another action). To ensure that, at each interaction, there was at least one decision to take between the human and the robot, we added to the objects to scan and store a red video tape box. The human and the robot both had a red video tape box in the same placement as their stacks of cubes. At the end of the task, when all the cubes were scanned and stored (and so both agents were available), only one of the two video tape boxes (the one of the human or the one of the robot) needed to be put on a scan area. Then, similarly to what happens with the cubes, the robot scanned the video tape box. Finally, since the video tape box was red, it needed to be stored in a red box either by the human or the robot.
- *Distraction task:* We noticed that some subjects tried not to miss any action of the robot (they stayed in front of the robot each time there was a cube to scan and they hurried in the places where they could not see the robot). Consequently, there was not missing knowledge during the task for these subjects. To ensure that all subjects missed some actions of the robot, at one predefined point of the task, the experimenter asked the subject to leave the task for a while to perform another task consisting in building with Lego bricks a construction shown in a picture. Once the construction was achieved,

the subject was free to go back to the main task.

C. The Conditions

As said before, the aim of this study was to compare two versions of a system allowing us to execute Shared Plans during human-robot Joint Action:

- *The Reference System (RS)*: In this system, the robot computes a Shared Plan where all decisions are taken in advance (who should do which action with which specific object). It means that for the task previously described, the robot will decide at the beginning of the interaction who will store which red objects and which scan area to use for each cube. However, if, at one point, the human does not act as planned, the robot will be able to compute a new plan and adapt. Moreover, in this system the robot does not compute the mental state of its partner concerning the Shared Plan. Consequently, in this task, the robot will not be able to identify whenever the human is not aware of a scan action and that this knowledge is important for the rest of the task. For this system, we choose to use two different modes corresponding to the different solution that this system allows concerning communication:
 - **SILENT**: In this mode the robot does not inform the human either about the choices it made or the information the human misses. The only possible verbalizations are whenever the situation is blocked by an action the human should perform and is not performing.
 - **VERBOSE**: In this mode the robot informs the human about all decisions it made (it informs before performing each action and each time the human has an action to perform). Moreover, the robot informs the human about all action he misses.
- *The New System (NS)*: In this system, the robot estimates the mental states of the human concerning the Shared Plan. It is able to identify whenever the human misses an information concerning an object which has been scanned and to communicate about this information. Moreover, the robot postpones some of the decisions concerning the Shared Plan to execution. For the task of the study, the robot postpones the decisions concerning who should store the red objects and which scan area to use. Then, there are two possibilities to take the decisions on who should store a red object:
 - **NEGOTIATION**: the robot can ask to the human if he wants to store an object and then it acts according to the answer.
 - **ADAPTATION**: the robot can wait a few time to see if the human executes the action, and, if not, it executes it.

An example of Shared Plans computed by both systems can be found in Fig. 3. As the Shared Plan for the task of the study is quite big, we present here a minimal example where there is only one red cube to deal with (which is initially located in the human stack).

D. Methodology

21 subjects took part in the study (8 women and 13 men). The participants ranged in age from 21 to 50 (mean = 27.38 y, sd = 7.07) and had normal or corrected-to-normal vision. They were all fluent in french. Each subject of the study had to interact with the robot to achieve the task previously described, and in the four conditions described previously. The order in which they were confronted to the different conditions was randomized. There were four different configurations for the stacks of the human and the robot. The attribution of each configuration to a condition was also randomized for each participant. At their arrival, the participants were introduced to the robot and the environment of the study by the experimenter. Then, participants were asked to read instructions explaining the task and its constraints. The experimenter checked the good understanding of the instructions and showed the placements of the different objects of the task. The participants were then asked to perform a quick familiarization task. In this task, the human and the robot had only one cube in their stacks (a blue for the human and a green for the robot). They had to put them in the scan areas, scan them and then store them in the appropriate boxes. There was no video tape box in the familiarization task. After each interaction with the robot (for each condition), the participants were asked to fill the questionnaire presented in this paper in order to evaluate their feeling concerning the robot and the interaction. In addition to this questionnaire, after each interaction with the robot (including the familiarization task), we asked participants to answer a small yes/no questionnaire. This questionnaire contains general questions about what happened during the interaction (e.g. "Do you think all the cubes have been scanned?"). The aim of this questionnaire was to remind the key points of the interaction to the subjects.

We formulate two hypotheses for this study:

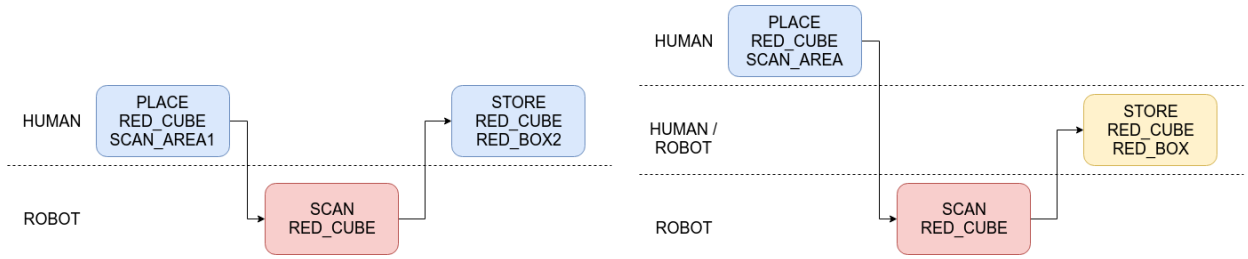
- **Hypothesis 1**: The new system will be preferred to the old one by the users.
- **Hypothesis 2**: For the new system, the negotiation mode will be preferred by the user to the adaptation one. Indeed, even if in simulation better results were found for the adaptation mode in [8], we believe that naive users will be more comfortable with a robot asking whenever there is a choice.

The entire experiment lasted between 45 minutes and one hour per participant.

V. RESULTS

A. Questionnaire

We calculated Cronbach's alpha [23] for each dimension of the questionnaire. These values can be found in Tab. II. To consider that the internal consistency reliability of a dimension is correct, alpha should be of 0.7 or higher. All the different dimensions have a cronbach's alpha greater than 0.7 except the Verbal dimension. It appears that these items do not all measure the same attribute. Therefore, this dimension will have to be reconsidered on a more suitable task to



(a) Shared Plan for the Reference System (RS), all actions are already allocated to an agent and all objects are chosen when the plan is built. (b) Shared Plan for the New System (NS), some decisions are deferred to last moment in order to provide more latitude to the human: the allocation of some actions and the choice of some objects or places

Fig. 3. Two different Shared Plans for the two systems compared in the study. The full shared plan can not be drawn here since it contains dozens of actions. We illustrate here a "minimal" example of a shared plans which concerns only one red cube situated in the stack close to the human.

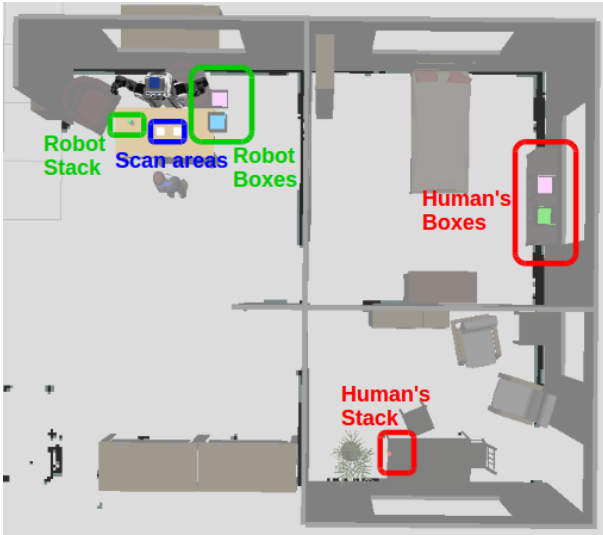


Fig. 4. Set up of the task used during the user study. The human and the robot had to bring the colored cubes from their respective stacks to the scan areas. Then the robot had to scan the cubes and finally the cubes had to be stored in the box corresponding to their colors. The areas in red in the figure can be reach only by the human, the ones in green only by the robot and the one in blue by both agents.

TABLE II
CRONBACH'S ALPHA FOR THE DIFFERENT DIMENSIONS OF THE QUESTIONNAIRE.

Dimension	Cronbach's alpha
Collaboration	0.86
Interaction	0.92
Robot perception	0.88
Verbal	0.6
Acting	0.87

Note: An alpha of 0.7 and higher means the dimension has an acceptable internal consistency reliability.

build a robust and validated questionnaire. However, some participants found that the fact that the robot was speaking a lot was comforting because they did not have to take decisions or to interpret robot actions. They also point out the fact that, even if they found it comforting the first time,

they would quickly find it "annoying" if they had to interact with the robot several times in this mode. This illustrates the difficulty of correctly evaluating the pertinence of verbal interventions of a robot by naive subjects on a short-term study. We will discuss this point in Sec. VI.

B. Results of the study

All the scores obtained for the different conditions can be found in Tab. III.

a) *Comparison of the two systems:* We first compared the Reference System (RS) to the New System (NS) in order to test the first hypothesis. We compared the two systems looking at the total score of the questionnaire and each dimension individually. We applied student T-tests when the data were normally distributed and Wilcoxon tests when the data were not normally distributed. The obtained results can be found in the first column of Tab. IV. We can see that for the total of the questionnaire and for all dimensions except the Verbal one, the new system has been evaluated significantly better than the reference system ($p < 0.05$). Consequently, we can consider the first hypothesis as validated. The difference was particularly visible for the *Acting* dimension of the questionnaire ($p \simeq 0.003$). It shows that the algorithms developed for the robot to be able to take the appropriate decisions at the right time during Shared Plan achievement have been appreciated by the subjects.

b) *Comparison of the negotiation and adaptation modes:* We then compared the negotiation and the adaptation modes of the new system in order to validate the second hypothesis. We compared the two modes looking at the total score of the questionnaire and each dimension individually. We applied student T-tests when the data were normally distributed and Wilcoxon tests when the data were not normally distributed. The obtained results can be found in the second column of Tab. IV. Here we can see that, even if the means of the negotiation mode are higher than the ones of the adaptation mode, no significant difference was found except for the Verbal dimension. Given that the internal consistency reliability of this dimension of the questionnaire is not acceptable, we cannot conclude about these particular results. Therefore, the second hypothesis is not validated: no difference has been found between the negotiation mode and the adaptation mode.

TABLE III
RESULTS FOR EACH DIMENSION OF THE QUESTIONNAIRE

Dimension	SILENT	VERBOSE	RS	NEGOTIATION	ADAPTATION	NS
Collaboration (/500)	301.29 ± 99.91	316.95 ± 101.54	309.12 ± 99.8	366.01 ± 98.43	357.43 ± 91.49	361.76 ± 93.96
Interaction (/500)	344.29 ± 98.23	355.95 ± 110.7	350.12 ± 103.53	383.52 ± 87.17	378.33 ± 99.95	380.93 ± 92.67
Robot Perception (/800)	382.05 ± 121.71	437.38 ± 126.41	409.71 ± 125.72	472.57 ± 143.79	457.67 ± 143.05	465.12 ± 141.86
Verbal (/300)	218.48 ± 53.83	227.29 ± 51.72	222.88 ± 52.33	254.43 ± 40.64	209.05 ± 63.93	231.74 ± 57.68
Acting (/300)	205.48 ± 61.76	200.86 ± 62.92	203.17 ± 61.62	238.33 ± 60.86	236.81 ± 41.43	237.57 ± 51.43
Total (/500)	318.19 ± 77.64	331.97 ± 81.2	325.08 ± 61.9	373.25 ± 74.66	352.98 ± 73.47	363.11 ± 59.09

Note: The RS and the NS columns are the means respectively for the Reference System and the New System. The presented results are the means for all subjects and their associated standard deviation. . The scores of the dimensions correspond to the addition of the scores of each question of the dimension (100 scale). The Total line is calculated by adding the scores of all dimensions previously harmonized in a 100 scale.

TABLE IV
P-VALUES FROM STUDENT T-TESTS AND WILCOXON TESTS (W).

Dimension	RS/NS	NEGO/ADAPT
Total	p = 1.325e-3*	p = 0.3108
Collaboration	p = 4.321e-3*	(W) p = 0.434
Interaction	p = 3.605e-2*	(W) p = 0.2371
Robot perception	(W) p = 2.080e-2*	(W) p = 0.5314
Verbal	p = 0.3075	(W) p = 8.966e-3*
Acting	p = 3.537e-3*	(W) p = 0.2219

Note: The first column corresponds to the comparison of the Reference System (RS) and the New System (NS). The second column corresponds to the comparison of the negotiation and the adaptation modes of the new system.

VI. DISCUSSION AND CONCLUSIONS

We have presented in this paper the PerDITA questionnaire, specifically designed to evaluate high level decision abilities of a robot during physical Joint Action with a human. Our aim was to propose this questionnaire as a preliminary work in order to build a strong and validated scale to measure the pertinence of robotic decisions on a Joint Action context. Following the results of this first study and correlation analyzes carried out, the questionnaire was modified with, especially the Perception of the Robot dimension which was replaced by a dimension on the Perceived Competence of the Robot. Indeed, it seemed more relevant to focus on this aspect in order to more accurately evaluate the pertinence of robot decisions, considering that the perception can be measured by the Godspeed questionnaire series [2]. This new version of the PerDITA questionnaire will be used in future studies and it remains to keep working on and test it on other tasks and contexts.

The pilot study presented here has also given to us the opportunity to get more insights on experiments with naive subjects. Indeed, there is a real difficulty to correctly evaluate technical contributions to decision for human-robot interaction due to several reasons:

- one of the first challenge is to find a task for the human and the robot to perform together. Indeed, this task should be sufficiently interesting to permit the

evaluation of the system but not too complex if we do not want the subject to focus too much on the task rather than on the robot behavior. Moreover, the task should be adapted to the current perception and manipulation abilities of robots.

- another difficulty when the robot interacts with a human is to isolate the decisional aspect from the other robot abilities. Indeed, we figured out that, because they are not used to interact with robots, it is quite hard for the subjects to distinguish two different behaviors of the robot (even if the difference seems huge to roboticists) because they have too much things to observe in the robot behavior.
- as the implementation of the robot decisional abilities usually relies on the other components of the robotics architecture (e.g. perception, manipulation), it is complicated and time consuming to obtain a global system sufficiently robust for a user study.
- finally, as the subjects during the pilot studies are naive, we can question the obtained results. Indeed, we have experimented the fact that it is difficult, in a one session, while trying to avoid introducing heavy biases, to ask naive users to distinguish between a one-shot use of a robot and its potential daily use [24]. Then "annoying", "repetitive", "intrusive", "delayed", "lacking of fluidity", "superfluous" behaviors will be certainly more severely evaluated. For the time being, the users are basically happy to "play" with the robot. This was perhaps enforced by the fact that in all the versions that we have proposed to the users, since the robot observes correctly the state and produces valid plans, the tasks is always finally achieved. Some UX models take into account the temporal aspect of the interaction[25], however, it usually implies to make long term user studies which are not easily feasible in our research context. One challenge would be to find a way to evaluate this aspect without a need of a long term study, maybe by performing user studies with subjects who are not naive but more used to robotics systems.

ACKNOWLEDGEMENTS

This research was supported by the French National Research Agency (ANR) grant reference ANR-16-CE33-0017

REFERENCES

- [1] N. Sebanz, H. Bekkering, and G. Knoblich, "Joint action: bodies and minds moving together," *Trends in cognitive sciences*, vol. 10, no. 2, pp. 70–76, 2006.
- [2] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International journal of social robotics*, vol. 1, no. 1, pp. 71–81, 2009.
- [3] M. Heerink, B. Krose, V. Evers, and B. Wielinga, "Measuring acceptance of an assistive social robot: a suggested toolkit," in *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*. IEEE, 2009, pp. 528–533.
- [4] M. Heerink, B. Kröse, V. Evers, and B. Wielinga, "Relating conversational expressiveness to social presence and acceptance of an assistive social robot," *Virtual reality*, vol. 14, no. 1, pp. 77–84, 2010.
- [5] K. Fischer, L. C. Jensen, S.-D. Suvei, and L. Bodenhausen, "Between legibility and contact: The role of gaze in robot approach," in *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*. IEEE, 2016, pp. 646–651.
- [6] S. Lemaignan, M. Warnier, E. A. Sisbot, A. Clodic, and R. Alami, "Artificial cognition for social human-robot interaction: An implementation," *Artif. Intell.*, vol. 247, pp. 45–69, 2017. [Online]. Available: <https://doi.org/10.1016/j.artint.2016.07.002>
- [7] S. Devin and R. Alami, "An implemented theory of mind to improve human-robot shared plans execution," in *Human-Robot Interaction (HRI), 2016 11th ACM/IEEE International Conference on*. IEEE, 2016, pp. 319–326.
- [8] S. Devin, A. Clodic, and R. Alami, "About decisions during human-robot shared plan achievement: Who should act and how?" in *ICSR'17*. to appear, 2018.
- [9] G. Hoffman, "Evaluating fluency in human-robot collaboration," in *International conference on human-robot interaction (HRI), workshop on human robot collaboration*, vol. 381, 2013, pp. 1–8.
- [10] M. C. Gombolay, R. A. Gutierrez, S. G. Clarke, G. F. Sturla, and J. A. Shah, "Decision-making authority, team efficiency and human worker satisfaction in mixed human-robot teams," *Autonomous Robots*, 2015.
- [11] A. D. Dragan, S. Bauman, J. Forlizzi, and S. S. Srinivasa, "Effects of robot motion on human-robot collaboration," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2015, pp. 51–58.
- [12] J. Brooke, "System usability scale (sus): a quick-and-dirty method of system evaluation user information," *Reading, UK: Digital Equipment Co Ltd*, 1986.
- [13] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User acceptance of information technology: Toward a unified view," *MIS quarterly*, pp. 425–478, 2003.
- [14] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, "The robotic social attributes scale (RoSAS)," in *Proceedings of the ACM/IEEE International Conference of Human-Robot Interaction*. ACM, 2017, p. 254262.
- [15] M. Hassenzahl and N. Tractinsky, "User experience—a research agenda," *Behaviour & information technology*, vol. 25, no. 2, pp. 91–97, 2006.
- [16] S. Mahlke, "User experience of interaction with technical systems," 2008.
- [17] A. Dillon, "User acceptance of information technology," in *Encyclopedia of human factors and ergonomics*. London: Taylor and Francis, 2001.
- [18] M. Hassenzahl, "The thing and i: understanding the relationship between user and product," in *Funology*. Springer, 2003, pp. 31–42.
- [19] C. Lallemand, V. Koenig, G. Gronier, and R. Martin, "Création et validation d'une version française du questionnaire attrakdiff pour l'évaluation de l'expérience utilisateur des systèmes interactifs," *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, vol. 65, no. 5, pp. 239–252, 2015.
- [20] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25(1), p. 4959, 1994.
- [21] J. Broekens and W.-P. Brinkman, "Affectbutton: A method for reliable and valid affective self-report," *International Journal of Human-Computer Studies*, vol. 71, no. 6, pp. 641–667, 2013.
- [22] M. Fiore, A. Clodic, and R. Alami, "On planning and task achievement modalities for human-robot collaboration," in *Experimental Robotics*. Springer, 2016, pp. 293–306.
- [23] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *psychometrika*, vol. 16, no. 3, pp. 297–334, 1951.
- [24] B. Kuhnert, M. Ragni, and F. Lindner, "The gap between human's attitude towards robots in general and human's expectation of an ideal everyday life robot," in *Robot and Human Interactive Communication (RO-MAN), 2017 26th IEEE International Symposium on*. IEEE, 2017, pp. 1102–1107.
- [25] H. K. Kim, S. H. Han, J. Park, and W. Park, "How user experience changes over time: A case study of social network services," *Human Factors and Ergonomics in Manufacturing & Service Industries*, vol. 25, no. 6, pp. 659–673, 2015.