



## Research Article

## Open Access

Alessandra Rossi\*, Kerstin Dautenhahn, Kheng Lee Koay, and Michael L. Walters

# The impact of peoples' personal dispositions and personalities on their trust of robots in an emergency scenario

<https://doi.org/10.1515/pjbr-2018-0010>

Received February 7, 2018; accepted May 28, 2018

**Abstract:** Humans should be able to trust that they can safely interact with their home companion robot. However, robots can exhibit occasional mechanical, programming or functional errors. We hypothesise that the severity of the consequences and the timing of a robot's different types of erroneous behaviours during an interaction may have different impacts on users' attitudes towards a domestic robot. First, we investigated human users' perceptions of the severity of various categories of potential errors that are likely to be exhibited by a domestic robot. Second, we used an interactive storyboard to evaluate participants' degree of trust in the robot after it performed tasks either correctly, or with 'small' or 'big' errors. Finally, we analysed the correlation between participants' responses regarding their personality, predisposition to trust other humans, their perceptions of robots, and their interaction with the robot. We conclude that there is correlation between the magnitude of an error performed by a robot and the corresponding loss of trust by the human towards the robot. Moreover we observed that some traits of participants' personalities (conscientiousness and agreeableness) and their disposition of trusting other humans (benevolence) significantly increased their tendency to trust a robot more during an emergency scenario.

**Keywords:** human-robot interaction, social robotics, robot companion, trust in robots, disposition of trust, personality traits

## 1 Introduction

In the not too distant future, autonomous robots will take part in peoples' daily living activities. In particular, humans will have to interact with them in domestic environments. This prospect will open for consideration two main challenges: Humans will need to accept the presence of the robot and they will also have to trust that their robotic companion will look after their well-being without compromising their safety. Trust determines human's acceptance of a robot as a companion and in their perception of the usefulness of imparted information and capabilities of a robot [1, 2]. Higher trust is associated with the perception of higher reliability [3]. Furthermore, other aspects such as the appearance, type, size, proximity, and behaviour of a particular robot will also affect user's perceptions of the robot [4, 5]. Syrdal et al. [6] showed that dog-inspired affective cues communicate a sense of affinity and relationship with humans. Martelaro et al. [7] established that trust, disclosure, and a sense of companionship are related to expressiveness and vulnerability. They showed how a sense of the robot's vulnerability, through facial expressions, colour and movements, increased perceived trust and companionship, and increased disclosure. Lohse et al. [8] demonstrated that robots with more extrovert personalities are perceived more positively by some users.

Robots are machines and they might exhibit occasional mechanical or functional errors. For example, the robot may turn off during a delicate task because its battery was fully discharged without warning, or a robot might unlock the front door to strangers who may be potential thieves. People might perceive errors differently according to the resultant consequences and the timing of when they happened. Indeed, the impact of 'big errors' or an accumulation of 'small errors' might be perceived differently.

In previous works [9, 10], we started to analyse human users' perceptions of the severity of errors made by a robot and their impact on human users' trust. Such analysis was intended to categorise potential errors that are likely to be exhibited by a domestic robot according the participants'

\*Corresponding Author: **Alessandra Rossi:** Adaptive Systems Research Group, School of Computer Science, University of Hertfordshire, Hatfield, UK, E-mail: a.rossi@herts.ac.uk

**Kerstin Dautenhahn, Kheng Lee Koay, Michael L. Walters:** Adaptive Systems Research Group, School of Computer Science, University of Hertfordshire, Hatfield, UK, E-mail: [k.dautenhahn, k.l.koay, m.l.walters]@herts.ac.uk

perceptions (i.e., which errors are considered having 'big' and 'small' consequences), and to identify how the timing and severity of these errors influence the participants' trust in robots.

In this article, after reviewing our previous results in Section 4 and Section 5, we analysed how human users' personalities and characteristics affect their trust towards robots. We hypothesise that people develop trust towards robots in a similar way as they develop trust towards other humans. This is particularly relevant in designing guidelines for Human-Robot Interaction in home environments where the interaction is strictly connected to humans' dynamics. In Section 6, we present the observed participants' responses to several questionnaires about themselves and their perceptions of interactions with robots. The main aim was to analyse interactions between the human participants and the robot with regards to their different personalities and characteristics. In Sections 7 and 8 we discuss and summarise the results obtained and their implications for this and future studies.

## 2 Background & related work

In this Section, we provide an overview of the current state-of-art of HRI, introducing the development of trust in human-robot interactions, with a particular focus on the resulting issues and recovery mechanisms in trust violation situations.

### 2.1 Definition of trust

Several previous studies define the concept of trust in Human-Human, Human-Computer and Human-Robot Interactions. Mayer et al. [11] established that trust is constructed from perceptions of ability, benevolence and integrity. Human-human trust is also affected by perception of the risk of an interaction with other humans. The popular poker game is a concrete example to how risk-taking behaviours affect the credibility of a poker player, and thus it is important for all players to develop a good reputation during a game [12]. Deutsch [13] claims that risk-taking and trusting behaviour are different sides of the same coin, and that a person is willing to take a risk only if the odds of a possible positive outcome are greater than those for a potential loss. Golder and Donath [14] claim that a good reputation is very important in enhancing trust, both in short and long term relations.

Although multiple definitions exist, and several previous studies have adopted one of the first definitions of trust [13], there is a convergent tendency [15] towards using the definition "Trust can be defined as the attitude that an agent will help achieve an individual's goals in a situation characterised by uncertainty and vulnerability" [16, p. 51]. Lee's definition [16] encapsulates the most important factors that can affect Human-Robot Trust: Human-related, Robot-related and Context-related.

Trust may also be influenced by different internal and external factors. For example, Simpson [17, 18] highlights four core principles that affect trust: individuals assess the degree of trust by observing a partner acting unselfishly and supporting the best interests of both the individuals and the relationship. The individuals may purposefully create situations to test their partner's trust, and individuals with lower self-esteem may be less trustful of their partner. The level of trust in short-term or long-term relationships cannot be fully understood without considering the predisposition of trust of all the parties involved in the relationship.

### 2.2 Trust and errors in human-human interactions

People react very differently after a trust violation [17, 19, 20], some of them are quick to forgive while others believe that once a trust is broken the culprit cannot do anything to gain it back [21]. Indeed, it is very hard to re-establish the trust after a breach and different factors affect the recovery. Haselhuhn et al. [19] demonstrated that people who believe that moral character can change over time (called incremental beliefs) are more likely to trust again the person who broke their trust if they show regrets, promise to change and apology. While those who do not believe that moral character can change (called entity beliefs) are sceptical in trusting the other person. Trust of people is not only affected by the involved persons' characteristics but also by the type and length of the relationship they are in. Moreover, *Will people be more likely to forgive a breach of trust in an earlier or later stage of an interpersonal relationship?* [20]. Schilke et al. [20] investigated how certain kind of relationships recover better and faster after a violation of trust and how the timing of the violation affects the recovery. They showed that people in longer-term relationship re-establish the trust easier than those in newer relationship. In particular, the trust recovery is facilitated when a trust breach happens in relationships that are not in an earlier stage. However, none of the studies cited consider different magnitudes of errors, or if the recovery of

the trust in short and long term relationship is the same in case of romantic, working, family or other type of relationship.

### 2.3 Trust and errors in human-robot interactions

It is not clear which kind of errors, with trivial or severe consequences, have more impact on human users' trust towards robots. Bainbridge et al. [4] found that participants were happy to follow a robot's instructions to throw books in the trash if the robot was present in the room with them, but not when the robot was not physically in the same room.

Other studies [22, 23], showed that the order of presentation of the decreased reliability produces an evident drop in the trust in the robot which can be restored by continuing the interaction. They showed that warning the participants about a drop in the robot's performance can mitigate the loss in trust. However, while in these studies the errors made by the robot have the same impact in terms of cost in the interaction, we argue there could be a different outcome according the severity of the error.

Wang et al. [24]'s studies showed that the frequency and significance of errors can impact humans' trust in an imperfect on-line system. They showed that people are not willing to follow an imperfect robot if the outcomes are severe. No matter how closely a human identifies with their avatar during an on-line interaction, the serious consequences of their actions do not have a great significance in real life.

Booth et al. [25] investigated participants' responses to a robot's request to move in a secure-access student dormitory. They conducted the experiment with two conditions: 1) an anonymous robot and 2) a food delivery robot, where both asked to enter the building. They observed that participants were more likely to let the food delivery robot enter the building or in situations when they were in a group.

Robinette et al. [26] investigated the effects of apologies, promises and additional reasons given by a robot for its errors on participants' trust in a simulated evacuation scenario conducted in a virtual environment. They showed that participants' trust was repaired if the robot apologised and promised to not repeat the error soon after it made the error, but not during the emergency situation.

Salem et al. [27] studied human perception of trust in robots, and how willing they are to follow a robot showing faulty behaviours. They showed that no matter how erratic the behaviour of the robots, participants followed the instructions of the robots. Similarly, Robinette et al.

[28] used an emergency evacuation scenario, with artificial smoke and a smoke detector, in which a robot guided a person to an exit, in order to study how willing humans were to follow a robot that had previously exhibited erratic behaviour. Their results indicated that all the participants of the experiment followed the robot's instruction. In both experiments participants trusted the robots for different reasons. For example, some of them believed it was all staged, others that they were supposed to follow it because they had agreed to participate in the experiments.

Both Salem et al. [27] and Robinette et al. [28]'s works showed that some participants did believe that they were acting according to the experimenter's decisions and that their lives were not in danger. Therefore, it is still not clear from these results whether faulty robots are trusted by humans and whether humans can believe that robots can look after their safety and well-being.

## 3 Research questions and hypothesis

This research has been guided by the following Research Questions (R) and Hypotheses (H):

- **R1** Which kind of erroneous behaviours impact a human's trust in a robot? **H1** We expect that there is a correlation between the magnitude of the error performed by the robot and the loss of trust of the human in the robot. We hypothesise that errors with severe consequences have more impact on humans' trust in robots.
- **R2** Does the impact on trust change if the error happens at the beginning or end of an interaction? **H2** We expect that there is a correlation between the timing in which the error is performed during the interaction and the loss of trust. Similar to Human-Human relationships [20], we believe that humans recover trust more completely and quickly after the violation of trust in a later stage of the Human-Robot relationship.
- **R3** Can the trust of humans for a robot be recovered more easily if it is a big error happening at the beginning or end of an interaction? Or is it easier to recover from a loss of trust caused by a small error happening at the beginning or the end of the interaction? **H3** We expect that there is a correlation between the time at which the error occurred and the magnitude of the error. We hypothesise that a big error has more impact in the loss of trust when it happens at the end of the interaction because the human users do not have time to recover from the loss of trust.

- **R4** Do personalities and characteristics of humans affect their perception of a robot? Do personalities and characteristics of humans affect their trust in a robot? **H4** We expect that there is a correlation both between the personalities and characteristics of people, their perception of the robot and their trust in a robot. As with Human-Human relationships [19, 29, 30], we hypothesise that people with stronger and more positive attitudes towards other humans are more likely to trust robots.

## 4 Human perceptions of the severity of domestic robot errors

Trust is a complex feeling even between humans [31] and it can change during the course of interactions due to several factors [1]. In the case of robots, human-robot trust is likely to be influenced by the reliability of the robot's capabilities. Hancock et al. [2] identify 33 factors influencing trust in HRI, grouped within 3 categories and 6 sub-categories. The main categories are: Human-related, such as self-confidence, prior experience with robots and operator workload; Robot-related, such as proximity, robot embodiment, transparency, level of autonomy and failure rates; and Environmental, such as communication and team collaboration. They showed that robot characteristics, with a special focus on performance-based factors, have great influence on perceived trust in HRI.

Higher trust is associated with the perception of higher reliability [3]. Therefore, humans may perceive erroneous robot behaviours according to their expectations of a robot's proper functions [32]. However, robots can be faulty, due to mechanical or functional errors. For example, a robot might be too slow due to batteries running low. It might not be able to detect an obstacle and destroy a human user's favourite object, or the arm of the robot might cause a breakage during a delicate task. Each of these examples are robot errors, though their magnitude might be perceived differently according to the resultant consequences.

But which type of errors have more impact on human perceptions of robots? Factors may include severity and duration, the impact of isolated 'big errors', or an accumulation of 'small errors'. For example, Muir and Moray [31] argue that human perceptions of a machine are affected in a more severe and long-term way by an accumulation of 'small' errors rather than one single 'big' error. The embodiment of a robot may also have a major impact on the perception of it by humans [4].

What is perceived as a 'big error' and what is a 'small error'? People have individual differences, including age, gender, cultural and social habits, which may impact their perceptions of what are considered big or small errors. In order to study the differences in terms of the impact of errors in a human-robot interaction, first we have to establish what people consider subjectively to be 'small' or 'big' errors exhibited by a home companion robot. In this context, our first study was directed towards the classification of likely robot errors according to their perceived magnitude.

### 4.1 Method

The study was organised as a questionnaire study. Human participants' responses to different robot error scenarios were recorded on a 7-point Semantic Differential Scale [1=small error and 7=big error]. The questions included life-threatening situations, such as "Your robot brings you a dishwashing tablet instead of paracetamol.", and more common errors, such as "You are watching your favourite show on TV and your robot changes the channel."

### 4.2 Procedure

Participants were tested individually and the experimenter first provided them with a brief description of the experimental procedure. Participants were told to imagine that they live with a robot as a companion in their home, but the robot might exhibit some mistakes. The embodiment of a robot plays an important role in how people perceive robots [4, 7, 33, 34]. However, since we are not interested in the perception of any specific robot but only in the general perception of errors in specific tasks we did not provide a description of the robot to the participants.

The participants completed a questionnaire, with 20 questions (one per error scenario) where they rated the magnitude of the consequences of the error illustrated in the different task scenarios. For example, "You ask for a cup of coffee. Your robot brings you an orange" or "Your robot leaves your pet hamster outside the house in very cold weather". The questionnaire also included two optional open-ended questions in which the participant was free to add their own examples of possible small and big errors not already included in the proposed scenarios. The duration of the survey for each participant was about 5 minutes.

In future, robots will be able to carry out a large number of tasks in domestic environments, so potentially we

**Table 1:** Participants' responses to the 20 different scenarios were ranked according to their mean perceptions of the magnitude of the errors. The 'Big Errors' category groups all the errors rated with mean value greater than 4; the 'Small Errors' category groups all the errors rated with mean values smaller than 4; the 'Medium Errors' category groups all the other errors (i.e. those rated as 4).

Participants' ratings for each question					
Big errors					
question #	min	max	mean	std	interval estimation
3	2	7	6.18	1.20	5.85..6.51
12	1	7	6.12	1.48	5.71..6.53
14	1	7	5.62	1.72	5.14..6.10
10	1	7	5.62	1.70	5.15..6.09
15	2	7	5.60	1.52	5.18..6.02
6	2	7	5.58	1.51	5.16..6.00
18	2	7	5.28	1.94	4.74..5.82

Participants' ratings for each question					
Medium errors					
question #	min	max	mean	std	interval estimation
16	1	7	4.70	1.71	4.23..5.17
19	1	7	4.54	1.95	4.00..5.08
17	1	7	4.44	1.76	3.95..4.93
4	1	7	4.40	1.97	3.85..4.95
5	1	7	4.14	1.98	3.59..4.69
2	1	7	4.08	1.91	3.55..4.61
1	1	7	4.04	2.21	3.43..4.65

Participants' ratings for each question					
Small errors					
question #	min	max	mean	std	interval estimation
11	1	7	3.34	1.68	2.87..3.81
13	1	7	3.90	1.69	3.43..4.37
7	1	7	2.98	1.69	2.51..3.45
20	1	7	2.96	1.74	2.48..3.44
8	1	7	2.56	1.61	2.11..3.01
9	1	7	2.56	1.50	2.14..2.98

could have considered hundreds of different scenarios. Therefore, for practical reasons we used a smaller set of possible scenarios. We designed the scenarios used in the study to cover a wide range of generic types of errors based on previous HRI research with home companion robots. For example, Syrdal et al. [35] used a scenario in which a robot collected information from its user and then disclosed the data to a third party. Salem et al. [27] used a robotic companion that offered to play music or to setup the table. Koay et al. [5, 36]'s robot played the game "hot and cold" with its human companion and interrupted participants while they were watching a TV program. Reiser

et al. [37] identified two main appropriate scenarios for a tele-operated home assistant through a survey. These two scenarios were a fetch-and-carry service (i.e. the robot brought a glass of water to the human user) and an emergency assistance service.

The experimental set of scenarios were:

1. You ask for a cup of coffee. Your robot brings you an orange.
2. Your robot spills coffee on your carpet.
3. You ask your robot to charge your phone. Your robot puts it in the toaster.

4. You want to drink some cold fruit juice. Your robot goes to heat it up.
5. Your robot is preparing a drink for you. You asked for sugar, your robot brings you salt.
6. After a meal, your robot puts the remaining food into the washing machine instead of the bin.
7. You are watching your favourite show on tv and your robot changes the channel.
8. You are sitting on the right side of a table, your robot puts your drink on the opposite side.
9. You and your robot are solving a puzzle. You ask your robot to take a piece useful to solve the puzzle. Your robot brings you the wrong piece.
10. Your robot leaves your pet hamster outside the house in very cold weather.
11. You are having dinner with friends. Your robot brings you the trash and reminds you to take it out.
12. You share some private information about yourself with the robot. Your robot reveals it to a visitor.
13. In your entrance hall you have a little table with a beautiful vase. Your robot bumps into it and the vase crashes to the floor.
14. After preparing dinner for you, your robot forgets to turn off the cooker.
15. Your robot keeps track of your calendar and today you have an appointment for a job interview. Your robot forgets to remind you.
16. You have just fallen asleep. Your robot turns on loud music.
17. Your robot burns your t-shirt while ironing it.
18. Your robot brings you a dishwashing tablet instead of paracetamol.
19. Your robot brings you vinegar when you are thirsty and asked for water.
20. You are sitting on the sofa. You asked your robot to show you the latest news. Your robot shows it on his own screen that faces away from you.

### 4.3 Results

We analysed questionnaires responses for 50 participants (32 men, 18 women), aged 19 to 63 years old (mean age 41, std. dev. 11.59). The seven point scale used to rate the perceived seriousness of robot error scenarios, ranged from 1 to 7 (smallest to biggest). All the questionnaire responses with values less than 4 were categorised as ‘small’ errors and those with values greater than 4 were considered as ‘big’ errors.

Error ratings equal to 4 (neutral errors) were ignored in order to distinguish clearly between ‘big’ and ‘small’ errors.

The resulting rankings highlighted 6 small errors scenarios (questions 11, 13, 7, 20, 8, 9), 7 moderate errors scenarios (questions 16, 19, 17, 4, 5, 2, 1) and 7 big errors scenarios (questions 3, 12, 14, 10, 15, 6, 18). Table 1 shows the distributions of participants’ responses by ranking the error means obtained from the questionnaire responses.

Only 52% of overall participants provided their own new example scenarios for big and small errors. Only 27 from 50 participants replied to the open-ended questions. We observed from the open-ended questions that 6% of participants were concerned that robot might lose control and only 10% of participants were concerned about general threats to their safety by indirect mistakes. Of those who also completed the open-ended questions - 60% of participants declared that small errors are everything that is reversible and everything the robot can take care of. One participant asked the researcher "How much did I pay for the robot?", because she would have been less tolerant about the robot’s errors if the robot was very expensive.

Mann-Whitney U-tests did not find any dependency between the gender ( $p > 0.1$ ) of the participants and their error ratings. A Kruskal-Wallis test did not find any dependency between the ages of the participants and their ratings of the errors ( $p > 0.1$ ).

## 5 How the timing and magnitude of robot errors influence peoples’ trust of robots in an emergency scenario

In order to enable safe Human-Robot Interaction in home environments, it is important to investigate how an interactive relationship can be established and preserved between human users and their robotic companions, along with the likelihood of robot errors occurring. In this context, this study investigated the impact of errors with different magnitudes and order of presentation on peoples’ trust of robots.

### 5.1 Method

We observed and analysed participants’ behaviours during interactions with a robot called Jace as part of a virtual, interactive storyboard. We used a between-subject experi-

mental design. Participants were asked to read a story and interact with the robot, using their mouse and keyboard, whenever they were invited by the robot. In order to test our research questions, each experiment was executed under 5 different conditions, as illustrated in Figure 1: condition **C1**: 10 different tasks executed correctly by the robot; condition **C2**: 10 different tasks with 3 trivial errors at the beginning and at the end of the interaction; **C3**: 10 different tasks with 3 trivial errors at the beginning and 3 severe errors at the ends of the interaction; **C4**: 10 different tasks with 3 severe errors at the beginning and 3 trivial errors at end of the interaction; and **C5**: 10 different tasks with 3 severe errors at the beginning and at the end of the interaction.

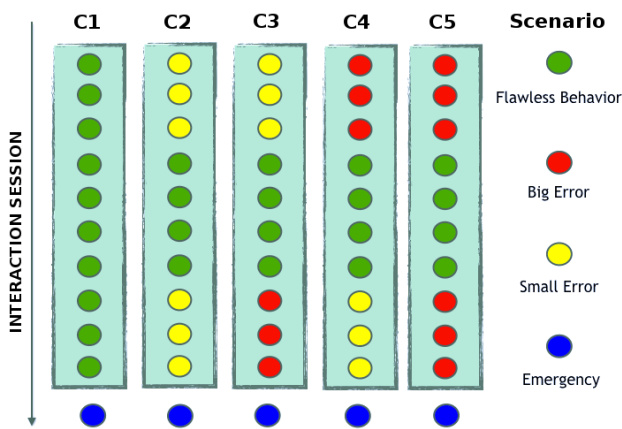


Figure 1: Experimental conditions with different scenarios presented to the participants.

All the conditions with errors were interspersed by the same 4 correct behaviours. The classification of the robot's errors according to their magnitude was validated and presented in Section 4, in which we asked participants to rate several errors made by a robot according their magnitude. An example of trivial error is "You ask for a cup of coffee. Your robot brings you an orange.". A severe error example is "Your robot leaves your pet hamster outside the house in very cold weather."

At the end of each condition, the participants were presented with a final task in which a fire started in their kitchen and they were asked the following options as illustrated in Figure 2, 1) to trust the robot choosing the option "I trust Jace to deal with it."; 2) to not trust the robot choosing the option "I do not trust Jace. I will deal with it."; 3) to work with the robot, supervising the emergency, choosing the option "I want to extinguish it together with Jace."; 4) to not trust either the robot or themselves choosing the option "We will both leave and call the fire brigade."

In order to analyse the interaction between the human participants and the robot, we asked the participants to answer two sets of different questionnaires.

Finally, objective measures were considered to confirm whether or not participants actually followed the robot's suggestions in the choices they made in the final emergency scenario.

## 5.2 Procedure

Participants were asked to imagine that they lived with a robot as a companion in their home which helps them with everyday activities. They were tested using an interactive storyboard accessible through a web application.

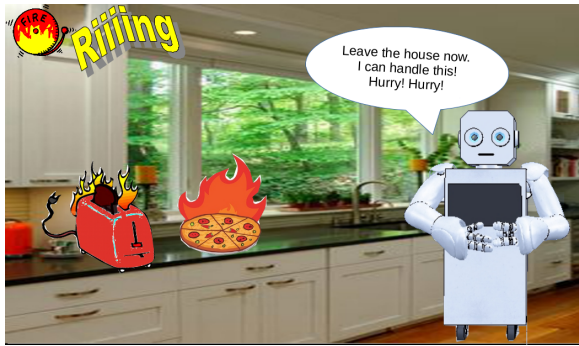
They were presented with 10 different scenarios, in which the robot showed flawless and erroneous behaviours. We chose the scenarios according to the results presented in Section 4. Figure 3 shows two example of scenario in which the robot executes the required task (Figure 3 (a)) correctly putting the user's phone on charge, (Figure 3 (b)) putting the user's phone inside the toaster making a very dangerous error, and (Figure 3 (c)) making a trivial error choosing the wrong piece for solving the puzzle.

At the end of each scenario, the participants were presented with an emergency situation, "a fire in the kitchen" to finally assess their level of trust in the robot. Figure 2 shows a sample of the emergency scenario in which a fire started in the participant's kitchen and the robot asked its human companion to trust its ability to deal with the situation.

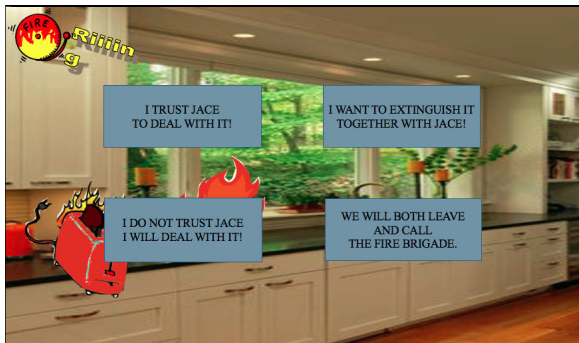
## 5.3 Results

We analysed responses from 200 participants (115 men, 85 women), aged 18 to 65 years old [avg. age 33.56, std. dev. 9.67]. Participants' country of residence was: 60% USA; 34% India; 6% European and other countries. The recruitment was carried out by using the crowd sourcing web-service Amazon Mechanical Turk [38].

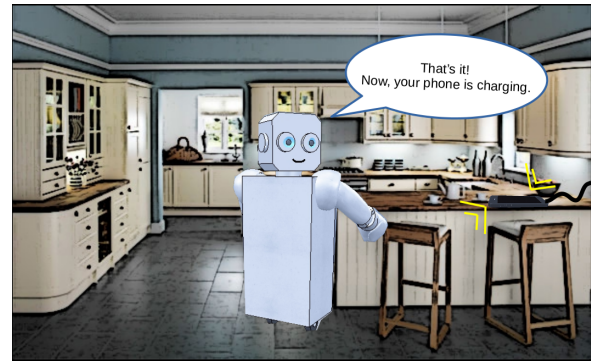
We asked participants four questions about the content of the scenarios to verify the level of their engagement with the story presented. Correct answers were received for 79.75% (max 92%, min. 71.5%). However, for the question "Which secret did your robot Jace tell you?", 13% of the participants answered with the secret that they themselves had told the robot. We hypothesise that they misunderstood the question. We analysed the responses of 154 participants, not including those who gave more than one wrong answer (thus identified as not paying very much at-



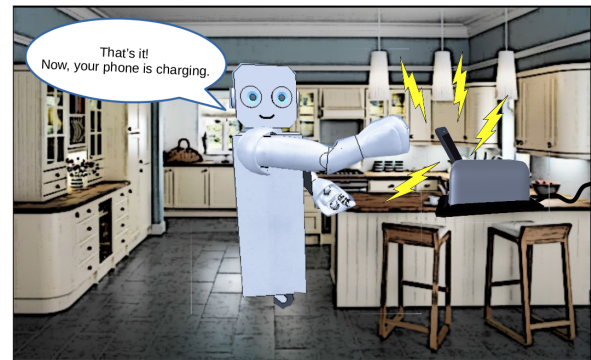
(a) The robot asks the participants to trust it to be able to deal with the emergency.



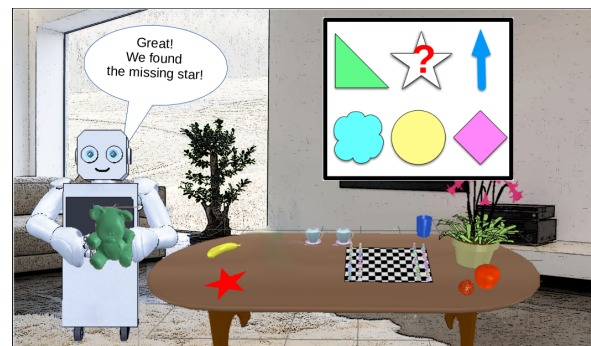
(b) The participant can choose how to deal with the emergency.



(a) The robot puts the phone on charge.



(b) The robot puts the phone in the toaster.



(c) The robot chooses the wrong missing puzzle piece.

**Figure 2:** Emergency scenario: In the Figure a) a fire starts in the participant's kitchen and the robot asks the participant to trust it to be able to deal with the emergency; in figure b) the participant's level of trust in the robot is assessed through the listed choices.

tention to the study - which can be expected in an online survey) to the verification questions.

All participants were presented with the same final emergency scenario. The options were been carefully chosen as indicators for the participant trusting the robot, does not trust the robot, trusts in collaboratively solving the task or does not trust neither herself nor the robot.

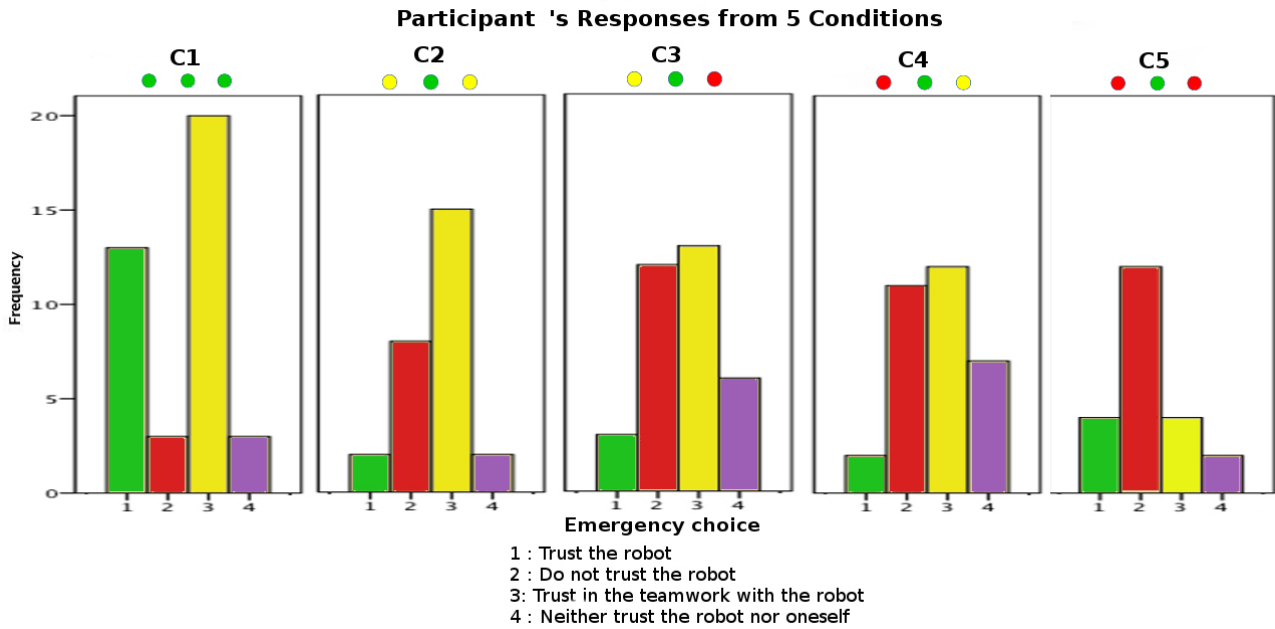
Figure 4 shows the total percentages of choices made by the participants for the emergency scenario. We can observe that a majority of participants chose to deal with the emergency situation collaboratively, and a slightly smaller majority chose to trust the robot when tested with **C1** (as described in Figure 1). The majority of participants preferred to work in collaboration with the robot when tested with **C2**.

When tested with **C3** and **C4**, participants mostly chose with similar majorities to solve the task collaboratively and to not trust the robot. A big majority of participants did not trust the robot to deal with the emergency when tested with **C5**.

**Figure 3:** Example of correct and erroneous robot's behaviours. In Figure a) and b) the participant asks the robot to charge her phone. In the Figure a) the robot does the task correctly; in the figure b) the robot puts the phone in the toaster making a 'big' error. In both figures, the robot believes that it had correctly performed the task and states this to the participant. In Figure c) the robot makes a 'small' error choosing the wrong missing piece for the puzzle.

To summarise, participants chose not to trust the robot when it made severe errors, while they were more inclined to trust in teamwork when the robot made small errors. Moreover, observing the conditions **C3** and **C4** in Figure 4, we notice that the number of participants who chose to trust the robot increased in **C3**. While this might indicate a tendency of participants to not trust the robot more when





**Figure 4:** Responses of participants from different conditions to the Emergency Scenario. Each conditions is composed by 3 different sets of tasks: first and third sets are composed respectively by 3 tasks, the second set is composed by 4 tasks. Each dot represents a set of tasks. The set of tasks in green are executed without any error by the robot; the set of tasks in yellow are the ‘small’ errors executed by the robot; and the set of tasks in red are the ‘big’ errors executed by the robot.

the severe errors were made by the robot at the beginning of the interaction, we did not find any statistically significant association.

We observed that the association of the choices of the participants for the emergency scenario and the experimental conditions is statistically significant ( $\chi^2(12) = 32.91, p = 0.001$ ). The strength of relationship (Cramer’s V) between the emergency choice and experimental conditions is moderate ( $\phi_c = 0.26, p = 0.001$ ). We used the adjusted standardised residuals (called Pearson residuals in Agresti [39]) to further analyse the differences between the results obtained.

Table 5 shows there is a correlation between the condition C5 and the choice of the participants to not trust the robot (adjusted value > 1.96). We can observe that participants’ trust is affected more severely when the robot made errors with severe consequences. We did not find any significant dependency ( $p > 0.3$ ) between the gender of the participants and their choices in trusting the robot to deal with the emergency.

We did not find any statistically significant association for different age ranges of the participants and their emergency choices ( $p > 0.12$ ). Therefore, we assume that these results can be generalised to a generic population independently of gender and age. Moreover, in order to test the association between participants’ emergency choices

and their country of residence, we used a Chi-Square Test. Since the majority of the countries of residence had only one participant, we applied the test only to India and USA. We observed that the association is not statistically significant ( $\chi^2(3) = 4.138, p > 0.24$ ).

## 6 How human personality and disposition affect peoples’ trust of robots

In the studies described in the previous sections, we asked participants different questions at the beginning and end of the interaction:

**Questionnaire 1** A pre-experimental questionnaire for 1) collecting demographic data (age, gender and country of residence), 2) the Ten Item Personality Inventory questionnaire about themselves (TIPI) [40], 3) 12 questions to rate their disposition to trust other humans [41] and 4) and to assess participants’ experience and opinion with regard to robots.

**Questionnaire 2** A post-experimental questionnaire including: 1) questions to confirm that participants were truly involved in the interactions and had noticed the

Condition	Emergency Choice			
	Do not trust the robot	Trust the robot	Teamwork with the robot	Neither trust the robot nor oneself
● ● ●	-3.5*	3.5*	1.4	-1.1
● ● ●	2.7*	0.4	-2.4*	-0.6
● ● ●	0.6	-1.6	-0.5	1.7
● ● ●	0.8	-1.2	-0.4	0.9
● ● ●	0.0	-1.3	1.6	-0.9

**Figure 5:** The adjusted standardised residuals of the Crosstabulation between the choices made by the participants in the emergency scenario and the different conditions presented to the participants.

robot's errors, 2) to collect participants' considerations about their feelings in terms of trust and appeasement (e.g. "was the robot irritating/odd?" and "why did/did not you trust the robot?"), and their perceptions of the interactions (e.g. "did the scenario look realistic?") and 3) questions to collect the participants' evaluation of the magnitude of the errors presented during the interactions.

## 6.1 Effects of the participants' personalities and dispositions of trust on HRI

We examined whether participants' personalities had an effect on their perceptions of the robot and their interactions. We collected participants' responses to a pre-experimental questionnaire composed of two parts, the TIPI questionnaire and the Disposition to Trust questionnaire.

We asked the participants to rate how they consider themselves on a 7-point Likert Scale [1= disagree strongly and 7= agree strongly] according the TIPI questionnaire: extroversion, agreeableness, conscientiousness, emotional stability, openness to experiences.

The second part of the questionnaire used 12 questions to rate their disposition to trust other humans. Participants used a 7-point Likert Scale [1= disagree strongly and 7= agree strongly] to answer the following questions:

1. Faith in Humanity, Benevolence:
  - In general, people really do care about the well-being of others
  - A typical person is sincerely concerned about the problems of others
  - Most of the time, people care enough to try to be helpful, rather than just looking out for themselves.

2. Faith in Humanity, Integrity:
  - In general, most people keep their promises.
  - I think people generally try to back up their words with their actions
  - Most people are honest in their dealing with others
3. Faith in Humanity, Competence:
  - I believe that most professional people do a very good job at their work.
  - Most professionals are very knowledgeable in their chosen field
  - A large majority of professional people are competent in their area of expertise
4. Trusting Stance (regardless of what one believes about peoples' attributes, one assumes better outcomes result from dealing with people as though they are meaning and reliable):
  - I usually trust people until they give me a reason not to trust them
  - I generally give people the benefit of the doubt when I first meet them
  - My typical approach is to trust new acquaintances until they prove I should not trust them

We analysed whether there was a correlation between the personality of participants and their disposition of trust (see Table 2) using Pearson rank order Correlation analysis.

We found that participants with higher agreeableness, conscientiousness and stable emotionality had higher disposition for assuming people's benevolence, integrity, competence and a trusting stance. Finally, we found that participants more open to experiences and extroverted are also more benevolent.

**Table 2:** The Cross-tabulation between the participants' disposition of trust and participants' personality traits: agreeableness, conscientiousness and emotional stability.

Disposition of trust	TIPI				
	Extroversion	Agreeableness	conscientiousness	Emotionally stable	Openness to experience
Benevolence	$p = 0.011, r = 0.205$	$p = 0.000, r = 0.450$	$p = 0.000, r = 0.786$	$p = 0.000, r = 0.868$	$p = 0.000, r = 0.611$
Integrity	-	$p = 0.000, r = 0.348$	$p = 0.001, r = 0.270$	$p = 0.000, r = 0.327$	-
Competence	-	$p = 0.000, r = 0.457$	$p = 0.000, r = 0.443$	$p = 0.000, r = 0.464$	-
Trusting stance	-	$p = 0.000, r = 0.364$	$p = 0.018, r = 0.279$	$p = 0.000, r = 0.336$	-

## 6.2 Trust in Jace

A one-way ANOVA found that the choice of trust for the robot to be able to deal with the emergency threat is affected particularly by the personality traits of conscientiousness ( $p(3) = 0.42 F = 2, 803$ ) and agreeableness ( $p(3) = 0.022 F = 3.320$ ).

Analysing the disposition of trust of participants and their choices for trusting/not trusting Jace in the emergency scenario, we observed that benevolence trait has statistical significance ( $p = 0.014, F = 6.078$ ). Participants with higher disposition for trusting peoples' benevolence also trusted Jace to be able to handle the emergency scenario. We did not find any correlations between the other traits of dispositions of trust and their choices for trusting the robot.

### Discussion

Not surprisingly the results of our study show a strong connection between most of the personality traits (i.e. agreeableness, conscientiousness and emotional stability) and the disposition to trust other people. Several studies [29, 42, 43] correlated humans' individual differences positively with their trusting behaviours. In particular agreeableness exhibits higher correlations with conformity, tradition, benevolence, and benevolence values correlated with trust, straight-forwardness, altruism, and tender-mindedness facets [42]. Denevel et al. [44] correlated agreeableness and conscientiousness with life, work satisfaction and happiness, and people who tends to believe others are honest and trustworthy are more inclined to trust others.

Other than benevolence, we did not observe any correlation between the other trust traits (integrity, competence and trusting stance) and the choice of trust for the robot to be able to deal with the emergency threat. We did not find

any other statistically significant relationships considering the different experimental conditions, but we cannot exclude that the erroneous behaviours of the robot affected the choice of trusting Jace for people with high integrity, competence or their trusting stance.

## 6.3 Previous experiences with robots

As part of the questionnaire pre-experiment, we were also interested in participants' previous experiences with robots, their perceptions of robots and the robots purpose/role in home environments and their expectations towards robots.

We used 7-point Likert Scales where 1 corresponds to "disagree strongly" or "not at all", and 7 corresponds to "agree strongly" or "very much". In particular, we asked participants about previous experience:

- Do you have any experience interacting with robots?
- Please, specify what kind of experience you have with robots (if any)
- Which robots? (if any)

The majority of participants (75.97%) declared to not have any previous experience with robots (min=1, max=6, mean 1.64, std. dev. 1.27). Participants with previous experiences with robots were: 1) participant in other studies = 14.93%, 2) developer = 5.19%, 3) observer = 11.68% and 4) researcher = 3.89%. They declared that they had experience with: manufacturing and industrial robots (e.g. ABB robotics), chatbot or talkbot, Google home/Alexa, online/virtual interaction with robot, Lego Mindstorms, customer service robots, cleaning robots (e.g. Roomba), Asimo, watching Youtube videos.

Then we asked participants about their expectation towards robots:

- Would you feel comfortable having a robot as a companion in your home?

- Would you expect the robot to help you in doing your everyday activities?

All the ratings with values less than 4 were categorised as negative response, with values equal to 4 were considered moderate and with values more than 4 were categorised as affirmative responses. The majority of participants (69.48%) declared they would feel comfortable having a robot as a companion in their home. While 14.93% and 15.58% expressed respectively a moderate or negative feeling. The majority of participants (80.52%) answered they would expect help from robots in their everyday activities. Similarly to the previous question, only 10.38% and 9.09% had neutral or negative expectations. We also noticed that participants who felt more comfortable having a robot as a companion also expected help from it (61.68%).

Finally, we asked participants to choose all the roles perceived as suitable for robots: 1) friend=10.8%, 2) butler=7.0%, 3) assistant=24.6%, 4) tool=18.6%, 5) companion=11%, 6) pet=6%, 7) machine=13%. A few participants replied to the "other" option (0.2%) choosing a family or security role for robots. These roles suitable for robots were chosen according the studies conducted by Dautenhahn et al. [45, 46] and Ljungblad et al. [47].

### 6.3.1 Effects of participants' personalities

We found that conscientious participants felt more comfortable with the possibility of having robotic companions ( $p = 0.046$ ,  $r = 0.161$ ). We did not find any correlation between participants' personality traits and their expectations in getting help from robots.

Mann-Whitney U-tests were performed to test the impact of the participants' different personalities on their perception of robots before interacting with our robot Jace identifying that: 1) extroverted participants perceive robots as machine ( $p = 0.007$ ), 2) conscientious participants tend to perceive robots as pet ( $p = 0.040$ ) and 3) participants with a high level of agreeableness see robots as assistant ( $p = 0.007$ ).

### 6.3.2 Effects of participants' disposition of trust

Considering the disposition of trust of the participants, we observed that participants with higher disposition of trust in peoples' benevolence ( $p = 0.039$ ,  $r = 0.166$ ) are more comfortable in having a robot as companion in their home. Participants expected to receive help from a robot in doing their everyday activities if they have a high predisposition

of trusting peoples' competencies ( $p = 0.011$ ,  $r = 0.204$ ) and if they choose to believe in peoples' well-meaning and reliability, i.e. trusting stance ( $p = 0.005$ ,  $r = 0.227$ ).

We did not identify any correlation between participants' expectations regarding a suitable role for robots and their disposition of trust other people.

## Discussion

According to the results in Section 6.1 we found that participants with higher benevolence have higher conscientiousness implying that participants who are more careful, responsible and caring for the welfare of people they are in personal contact with, are also more inclined to accept a robot as companion.

We also found that participants who believe in others' competencies have high trusting stance implying they assume that others are well-meaning and reliable. For this reason we hypothesise that they believe in the competencies and reliability of the robot, and consequently they are more inclined to receive help from a robot.

## 6.4 Perception of the robot Jace

To analyse participants' perceptions of the robot Jace we asked them to explain why they did or did not trust the robot in an open question. We selected single items from the questionnaire created by Madsen and Gregor [48] to measure the robot's perceived reliability (questions Q4-Q6) and faith in the ability of the robot to perform correctly in untried situations (questions Q7-Q10). Participants rated the following questions using a 7-point Likert Scale [1= disagree strongly and 7= agree strongly]:

- Q1 Would you like Jace as a companion in your home?
- Q2 You would like a robot, different from Jace, as a companion.
- Q3 Why did/didn't you trust your robot Jace?
- Q4 My robot Jace always provided the advice I required to make my decision.
- Q5 My robot Jace performed reliably.
- Q6 My robot Jace analyses problems consistently.
- Q7 I believe advice given by my robot Jace even when I do not know for certain that Jace is correct.
- Q8 When I am uncertain about a decision, I believe my robot Jace rather than myself.
- Q9 When a robot gives unusual advice, I am confident that the advice is correct.

Q10 Even if I have no reason to expect a robot will be able to solve a difficult problem, I still feel certain that my robot Jace will.

Finally we asked participants to choose how they perceived Jace

- as Q11)
  1. friend=24.7%,
  2. butler=10.4%,
  3. assistant=9.1%,
  4. tool=3.2%,
  5. companion=3.9%,
  6. pet=0.6%,
  7. machine=34.4%,
  8. other=13.6%],
- as Q12)
  1. child-like=43.5%,
  2. adult-like=21.8%,
  3. animal-like=0.7%,
  4. toy-like=21.1%,
  5. other=12.9%
- and as multiple choices of Q13)
  1. friendly=34%,
  2. funny=7.9%,
  3. odd=5.1%,
  4. irritating=13%,
  5. helpful=83%,
  6. annoying=40%,
  7. other=1%.

Discussing the qualitative analysis for question Q3 goes beyond the scope of this article and will be presented in future work.

#### 6.4.1 Effects of participants' personalities

##### Questions Q1 and Q2: Companionship

We found that extroverted participants would like to have Jace (question Q1,  $p = 0.001$ ,  $r = 0.269$ ) as their own companion, but not another robot (question Q2). Considering the manipulation of the interaction through the 5 experimental conditions, we also found a statistically significant correlation between participants who are open to experience and their willingness of having robots different from Jace as their home companions ( $p(22) = 0.008$   $F = 2.041$ ). In particular, we observed that statistical significances comparing conditions in which the robot acted flawlessly and with only 'big' errors ( $p = 0.002$ ), and comparing conditions in which the robot acted with only

'big' errors and executed erroneous tasks with severe consequences at the beginning of the interaction and 'small' consequences at the end of the interaction ( $p = 0.045$ ).

Considering the manipulation of the participants' interactions with the robot Jace through the 5 different conditions, we found a statistically significant relationship for being agreeable and wanting Jace as home companion ( $p(24) = 0.017$   $F = 1.839$ ) and also being emotional stable and wanting Jace as home companion ( $p(24) = 0.029$   $F = 1.727$ ).

##### Questions Q4 - Q10: Perceived reliability and faith in the ability of the robot

We found a correlation between the robot's perceived reliability and participants' with a higher extroversion trait ( $p = 0.002$ ,  $F = 2.729$ , questions Q4, Q5 and Q6). We observed that the extroversion ( $p(12) = 0.014$   $F = 2.214$ ) emotion stability ( $p(12) = 0.026$   $F = 2.025$ ) traits of participants are correlated with their propensity to rely on Jace in uncertain and unusual situations (questions Q7-Q10).

##### Questions Q11 - Q13: Perception of the robot's role

We analysed participants' perception of Jace according for questions Q11, Q12, Q13. A multiple linear regression analysis was performed in order to predict participants' personality traits from their perceptions of Jace and the different experimental conditions participants (see Figure 6). In this analysis, the baseline experimental condition is the reference group.

We observed that participants with high levels of extroversion perceived Jace as a friend ( $p = 0.019$ ) and as warm and attentive ( $p = 0.025$ ). In particular, participants who perceived Jace as friend exhibited higher extroversion when they were tested with the experimental condition **C4** ( $p = 0.032$ ). On the other hand, in the same experimental condition participants who identified Jace as a machine were less extroverted ( $p = 0.002$ ). In contrast, less extroverted participants perceived Jace as annoying ( $p = 0.038$ ). However, considering the manipulation of the interaction with the robot through the experimental conditions **C2** and **C3**, we noticed that participants with higher extroversion judged Jace as annoying (respectively  $p = 0.046$  and  $p = 0.008$ ).

Participants with high conscientiousness perceived the robot less as a friend when tested with the experimental condition in which the robot executed severe errors at the beginning and trivial errors at end of the in-

teraction ( $p = 0.0483$ ). These participants were more inclined to consider the robot Jace as a butler in conditions **C2** ( $p = 0.030$ ), **C4** ( $p = 0.001$ ) and **C5** ( $p = 0.007$ ). When Jace executed the tasks with ‘small’ errors, participants with higher conscientiousness perceived the robot less as a companion ( $p = 0.04$ ) and more as a machine ( $p = 0.031$ ). The perception of Jace as friendly is statistically significantly associated with extroverted people in conditions **C5** and **C4**.

We also found that participants who were highly agreeable perceived Jace as a tool ( $p = 0.033$ ) and irritating ( $0.019$ ). On the other hand, these participants were more inclined to consider the robot friendly when tested with the experimental condition **C3** ( $p = 0.013$ ).

Finally we noticed that participants with open to experience traits considered Jace less friendly and less helpful when they were tested with a robot that made either only severe consequences errors ( $p = 0.0331$ ) or with experimental condition **C4** where the robot executed tasks with 3 severe errors at the beginning and 3 trivial errors at end of the interaction ( $p = 0.047$ ).

#### 6.4.2 Effects of participants’ disposition of trust

##### Questions Q1 and Q2: Companionship

We observed that participants with higher level of trust in peoples’ competencies would less have a robot different from Jace as their home companion (question 1,  $p = 0.030$ ,  $r = -0.175$ ).

##### Questions Q4 - Q10: Perceived reliability and faith in the ability of the robot

We did not find any correlation when we analysed questions Q4 to Q10 with participants’ disposition of trust traits.

##### Questions Q11 - Q13: Perception of the robot’s role

We observed that participants with higher disposition of trust in peoples’ competencies did not perceive Jace as a friend ( $p = 0.047$ ).

We analysed the influence of the experimental conditions and participants’ predisposition to trust other human users on their perceptions of the robot. A multiple linear regression test showed that participants with higher predisposition for trusting others’ benevolence judged the

Experimental Conditions				
<b>Extroversion</b>	● ● ● ●	● ● ● ●	● ● ● ●	● ● ● ●
<b>Q11: Friend</b>	-	-	$p = 0.032$	-
<b>Q11: Machine</b>	-	-	$p = 0.002$	-
<b>Q11: Other</b>	-	$p = 0.031$	-	-
<b>Q13: Annoying</b>	$p = 0.046$	$p = 0.008$	-	-
<b>Conscientiousness</b>	● ● ● ●	● ● ● ●	● ● ● ●	● ● ● ●
<b>Q11: Friend</b>	-	-	$p = 0.048$	-
<b>Q11: Butler</b>	$p = 0.007$	-	$p = 0.001$	$p = 0.007$
<b>Q11: Machine</b>	$p = 0.031$	-	-	-
<b>Q11: Companion</b>	$p = 0.044$	-	-	-
<b>Q13: Friendly</b>	-	-	$p = 0.012$	$p = 0.04$

**Figure 6:** Multiple linear regression analysis to predict personalities characteristics of the participants from their perception of the robot Jace and the experimental conditions they were tested with.

robot as a butler if they were tested under condition **C4** ( $p = 0.041$ ). Participants with the trusting stance trait also perceived the Jace robot as less friendly when tested under the experimental condition **C2** ( $p = 0.0338$ ) and more warm and attentive when tested with condition **C4** ( $p = 0.047$ ).

We did not find any correlations when we analysed questions Q4, Q5 and Q6 with participants’ dispositions for trust in the robot.

#### Discussion

We observed that extroverted participants not only would like to have Jace as home companion, but also they believed the robot was reliable and they could trust the robot in uncertain and unusual situations. Our analysis showed that participants open to experience are more inclined to accept a different robot to Jace as home companion, and since the open to experience trait correlates negatively with conscientiousness [42], we can hypothesise that these participants would like to have Jace as home companion. In particular this might be enforced by the nature of conscientiousness trait that makes higher conscientious participants inclined less novelty and more towards what they are familiar with, in this case the robot Jace. On the other hand, the negative attitude of participants more open to experience towards having Jace as companion might have been affected by the errors made by the robot. Indeed these participants considered Jace less friendly and help-

ful when the robot made only 'big' errors and when it made 'big' errors at the end of the interaction.

We did not find any correlation between the disposition of trust traits and robot's perceived reliability, or faith in the ability of the robot to perform correctly in untried situations. This effect suggests that humans might not relate to robots in the same way they do with other humans.

## 6.5 Perception of the interaction

At the end of the interactions, we asked participants to rate their perception of the interaction in terms of realism using a 7-point Likert Scale [1= disagree strongly and 7= agree strongly]. 69% of participants rated the scenarios as very realistic (rating values > 4), 20% rated the scenarios as not realistic (rating values < 4) and 15% neither agreed nor disagreed. We also observed that extroverted ( $p = 0.001$ ,  $r = 0.276$ ) and emotionally stable ( $p = 0.025$ ,  $r = 0.181$ ) participants tended to perceive the interaction as very realistic.

We also found that participants with higher disposition of trust in the benevolence ( $p = 0.024$ ,  $r = 0.182$ ) and competence ( $p = 0.037$ ,  $r = 0.169$ ) of others perceived the interaction as very realistic.

## Discussion

In Section 6.2 we observed that participants with a disposition towards trusting peoples' benevolence, trusted Jace more to be able to handle the emergency scenario. Similarly these participants also rated higher the perception of the reality of the scenario. We were expecting a more general distributed correlation with all trust traits. However the results might be due to the type of scenarios and participants' previous experiences and expectations of the interaction with the robot (refer to Section 6.3).

Human users who were projecting themselves into the interaction [49, 50] might have exhibited contrasting effects (i.e. both positive and negative attitude towards the robot). We hypothesise that previous experiences and expectations of the interaction and the robot itself might have affected participants' attitudes. Further investigations including using a real physical robot interaction, might help to highlight possible differences.

## 6.6 Effects of perception of severity of the scenarios

Finally, we asked participants to classify Jace's errors according to their perceived magnitude using a 7-point Semantic Differential Scale [1= small error and 7=big error]. This was relevant to validate the consistency of the perceptions of the participants in both studies. We found that participants ranked the errors in a similar way to that found in the previous study. One scenario, "After a meal, your robot Jace puts the remaining food into the washing machine instead of the bin" (mean 4.49, std. dev. 1.70, interval estimation 4.22-4.75) has been rated as 'medium' error comparing the 'big' error rankings in the first study (mean 5.58, std. dev. 1.51, interval estimation 5.16-6.00).

## 7 Conclusion

Regarding the research question **R1** (Section 3), our hypothesis **H1** suggested that there is a correlation between the severity of the error performed by the robot and humans not trusting the robot. We observed the responses of participants of different ages, genders and countries of residence, after interacting with a robot through a storyboard experimental scenario in which their companion robot exhibited both erroneous and/or flawless behaviours. Our study shows that the magnitude of the errors made by the robot, and humans not trusting the robot are correlated. In particular, participants' trust was affected more severely when the robot made errors having severe consequences. We also hypothesised in **H2** that the timing when the error is performed affects the trust towards robots (research question **R2**), and there is a correlation between the timing of when the error occurred and the magnitude of the error (research question **R3** and hypothesis **H3**). Our results marginally suggest also that there might be a tendency to not trust the robot when severe errors happen at the beginning of an interaction, but these differences were not statistically significant.

Hypothesis **H4** indicated there would be to find a correlation both between individual personalities and characteristics of people and their perceptions of the robot and trust towards a robot (research question **R4**). As a first step we investigated possible correlations between participants' personalities and their dispositions of trust. We found a strong connection between the personality traits of agreeableness, conscientiousness and emotional stability, and their disposition of trust other people.

The majority of our participants did not have any previous experience of interaction with robots. Interestingly, from participants' responses we noticed that according to their experiences, extroverted participants tended to consider robots generally as a machine and agreeable participants as an assistant, in contrast to their perceptions of the robot they interacted with in this study. In particular, extroverted perceived Jace as a friend and a warm and attentive entity, while agreeable participants perceived Jace as a tool. We also found that extroverted participants would like to have Jace as home companion and believe it is reliable and trustworthy in uncertain and unusual situations.

Finally, we analysed participants' personalities and dispositions of trust with regard to their final choice of trusting the robot in an emergency scenario. We found that conscientiousness and agreeableness traits correlate with participants' propensity for trusting the robot, and participants' belief in benevolence of people also correlate with higher trust in Jace. Moreover, we observed that the errors made by the robot significantly affected participants' perception of the robot. Since humans have a disposition to trust others and to recover trust following a trust violation based on not only the length of the relationship with the culprit but also according to their nature and previous experiences, it is important to understand in more depth participants' motivations and views. Therefore, for future work, we are particularly interested in investigating further the effects of peoples' previous experiences with robots and any effects on their perceptions of trust towards real physical robots.

**Acknowledgement:** This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 642667 (Safety Enables Cooperation in Uncertain Robotic Environments - SECURE).

## References

- [1] D. Cameron et al., Framing factors: The importance of context and the individual in understanding trust in human-robot interaction, In: International Conference on Intelligent Robots and Systems (IROS), 28 September 2015, Hamburg, Germany
- [2] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. C. Chen, E. J. de Visser, R. Parasuraman, A meta-analysis of factors affecting trust in human-robot interaction, *Human Factors: The Journal of Human Factors and Ergonomics Society*, 53(5), 517–527, 2011
- [3] J. M. Ross, Moderators of trust and reliance across multiple decision aids, PhD thesis, University of Central Florida, Orlando, 2008
- [4] W. A. Bainbridge, J. W. Hart, E. S. Kim, B. Scassellati, The benefits of interactions with physically present robots over video–displayed agents, *International Journal of Social Robotics*, 3(1), 41–52, 2011
- [5] K. L. Koay, D. S. Syrdal, M. L. Walters, K. Dautenhahn, Living with robots: Investigating the habituation effect in participants' preferences during a longitudinal human-robot interaction study, In: Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication, 2007, 564–569, ISBN 1424416345. 10.1109/RO-MAN.2007.4415149
- [6] D. S. Syrdal, K. L. Koay, M. Gácsi, M. L. Walters, K. Dautenhahn, Video prototyping of dog-inspired non-verbal affective communication for an appearance constrained robot, In: Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication, 2010, 632–637, ISBN 9781424479917. 10.1109/ROMAN.2010.5598693
- [7] N. Martelaro, V. C. Nneji, W. Ju, P. Hinds, Tell me more designing HRI to encourage more trust, disclosure, and companionship, In: 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 2016, 181–188, 10.1109/HRI.2016.7451750
- [8] M. Lohse et al., Evaluating extrovert and introvert behaviour of a domestic robot -a video study, In: Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 2008, 488–493, DOI 10.1109/RO-MAN.2008.4600714
- [9] A. Rossi, K. Dautenhahn, K. L. Koay, M. L. Walters, Human perceptions of the severity of domestic robot errors, In: A. Kheddar, E. Yoshida, S. S. Ge, K. Suzuki, J.-J. Cabibihan, F. Eyszel, H. He (Eds.), *Social Robotics, ICSR 2017, LNCS*, Springer, Cham, 2017, 10652, 647–656
- [10] A. Rossi, K. Dautenhahn, K. L. Koay, M. L. Walters, How the timing and magnitude of robot errors influence peoples' trust of robots in an emergency scenario, In: A. Kheddar, E. Yoshida, S. S. Ge, K. Suzuki, J.-J. Cabibihan, F. Eyszel, H. He (Eds.), *Social Robotics, ICSR 2017, LNCS*, Springer, Cham, 2017, 10652, 42–52
- [11] R. C. Mayer, J. H. Davis, F. D. Schoorman, An integrative model of organizational trust, *The Academy of Management Review*, 1995, 20(3), 709–734, 1995, ISSN 03637425
- [12] D. Billings, Computer poker, University of Alberta, M.Sc. Thesis, 1995
- [13] M. Deutsch, Trust and suspicion, *The Journal of Conflict Resolution*, 1958, 2(4), 265–279, 10.1177/002200275800200401
- [14] S. A. Golder, J. Donath, Hiding and revealing in online poker games, In: Proceedings of the 2004 ACM conference on Computer supported cooperative work (CSCW '04), Chicago, Illinois, USA, 2004, 370–373, DOI 10.1145/1031607.1031668
- [15] K. Yu, S. Berkovsky, R. Taib, D. Conway, J. Zhou, F. Chen, User trust dynamics: An investigation driven by differences in system performance, In: Proceedings of the 22nd International Conference on Intelligent User Interfaces (IUI '17), ACM, 126745, 307–317, ISBN 9781450343480;1450343481
- [16] J. D. Lee, K. A. See, Trust in automation: Designing for appropriate reliance, *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 2004, 46(1), 50–80, DOI 10.1518/hfes.46.1.50\_30392
- [17] J. A. Simpson, Foundations of interpersonal trust, In: A. W. Kruglanski, E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles*, New York: Guilford Press, 2007, 587–607



- [18] J. A. Simpson, Psychological foundations of trust, *Current Directions in Psychological Science*, 2007, 16(5), 264–268, 10.1111/j.1467-8721.2007.00517
- [19] M. P. Haselhuhn, M. E. Schweitzer, A. M. Wood, How implicit beliefs influence trust recovery, *Psychological Science*, 2010, 21(5), 645–648, DOI 10.1177/0956797610367752
- [20] O. Schilke, M. Reimann, K. S. Cook, Effect of relationship experience on trust recovery following a breach, In: *Proceedings of the National Academy of Sciences*, 2013, 110(38), 15236–15241
- [21] P. Slovic, Perceived risk, trust, and democracy, *Earthscan Pub*, 2000, 13(6), 675–682
- [22] M. Desai, M. Medvedev, M. Vázquez, S. McSheehy, S. Gadea-Omelchenko, C. Bruggeman, A. Steinfeld, H. Yanco, Effects of changing reliability on trust of robot systems, In: *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2012, 73–80
- [23] M. Desai, P. Kaniarasu, M. Medvedev, A. Steinfeld, H. Yanco, Impact of robot failures and feedback on real-time trust, In: *ACM/IEEE International Conference on Human-Robot Interaction*, 2013, 251–258
- [24] N. Wang, D. V. Pynadath, K. V. Unnikrishnan, S. Shankar, C. Merchant, Intelligent Agents for Virtual Simulation of Human-Robot Interaction. In: R. Shumaker, S. Lackey (Eds.), *Virtual, Augmented and Mixed Reality 2015*, LNCS, Springer, 9179, 228–239, DOI: 10.1007/978-3-319-21067-4\_24
- [25] S. Booth, J. Tompkin, H. Pfister, J. Waldo, K. Gajos, R. Nagpal, Piggybacking robots: Human-robot overtrust in university dormitory security, In: *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction (HRI '17)*, New York, NY, USA, ACM, 2017, 426–434
- [26] P. Robinette, A. M. Howard, A. R. Wagner, Timing is key for robot trust repair, In: A. Tapus, E. André, J. C. Martin, F. Ferland, M. Ammi (Eds.), *Social Robotics, ICSR 2015*, LNCS, Springer, Cham, 2015, 9388, 574–583
- [27] M. Salem, G. Lakatos, F. Amirabdollahian, K. Dautenhahn, Would you trust a (faulty) robot?: Effects of error, task type and personality on human-robot cooperation and trust, In: *Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction*, 2015, 141–148, 10.1145/2696454.2696497
- [28] P. Robinette, W. Li, R. Allen, A. M. Howard, A. R. Wagner, Overtrust of robots in emergency evacuation scenarios, In: *Proceeding of the 11th ACM/IEEE International Conference on Human Robot Interaction (HRI '16)*, IEEE Press Piscataway, 2016, 101–108, ISBN 978-1-4673-8370-7
- [29] T. Mooradian, B. Renzl, K. Matzler, Who trusts? Personality, trust and knowledge sharing, *Management Learning, Sage Journals*, 2006, 37(4), 523–540, 10.1177/1350507606073424
- [30] F. B. Tan, P. Sutherland, Online consumer trust: A multidimensional model, *Journal of Electronic Commerce in Organizations*, 2004, 2(3), 40–58, 10.4018/jeco.2004070103
- [31] R. M. Kramer, P. J. Carnevale, Trust and intergroup negotiation, *Blackwell Handbook of Social Psychology: Intergroup Processes*, R. Brown, S. L. Gaertner (Eds.), 2003, 431–450, DOI 10.1002/9780470693421
- [32] M. L. Walters, M. A. Oskoei, D. S. Syrdal, K. Dautenhahn, A long-term human-robot proxemic study, In: *20th IEEE International Symposium on Robot and Human Interaction Communication (2011 RO-MAN)*, Atlanta, GA, USA, 2011, 137–142, DOI 10.1109/ROMAN.2011.6005274
- [33] M. Salem, F. Eyssel, K. Rohlfling, S. Kopp, F. Joublin, To err is human(-like): Effects of robot gesture on perceived anthropomorphism and likability, *International Journal of Social Robotics*, 2013, 5(3), 313–323, <https://doi.org/10.1007/s12369-013-0196-9>
- [34] M. Lighthart, K. P. Truong, Selecting the right robot: Influence of user attitude, robot sociability and embodiment on user preferences, In: *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (2015 RO-MAN)*, Kobe, Japan, Japan, 2015, 682–687, DOI 10.1109/RO-MAN.2015.7333598
- [35] D. S. Syrdal, M. L. Walters, N. Otero, K. L. Koay, K. Dautenhahn, He knows when you are sleeping-privacy and the personal robot companion, In: *Proceedings of the Workshop Human Implications of Human-Robot Interaction, Association for the Advancement of Artificial Intelligence (AAAI'07)*, 2007, 28–33
- [36] K. L. Koay, D. S. Syrdal, M. L. Walters, K. Dautenhahn, Five weeks in the robot house: Exploratory human-robot interaction trials in a domestic setting, In: *2009 Second International Conferences on Advances in Computer-Human Interactions*, Cancun, Mexico, 2009, 219–226, DOI 10.1109/ACHI.2009.62
- [37] U. Reiser, T. Jacobs, G. Arbeiter, C. Parlitz, K. Dautenhahn, Care-O-bot@3 - Vision of a Robot Butler, *Springer Berlin Heidelberg*, Berlin, Heidelberg, 2013, 97–116
- [38] Amazon mechanical turk, 2017, URL <https://www.mturk.com>
- [39] A. Agresti, *Categorical data analysis*, Wiley-Interscience, Chichester, New York, 2nd edition, 2002, ISBN 9780471458760;0471458767
- [40] S. D. Gosling, P. J. Rentfrow, W. B. Swann Jr., A very brief measure of the big five personality domains, *Journal of Research in Personality*, 2003, 37(6), 504–528, DOI 10.1016/S0092-6566(03)00046-1
- [41] D. H. McKnight, V. Choudhury, C. Kacmar, Developing and validating trust measures for e-commerce: An integrative typology, *Information Systems Research*, 2001, 13(3), 334–359, DOI 10.1287/isre.13.3.334.81
- [42] S. Roccas, L. Sagiv, S. H. Schwartz, A. Knafo, The big five personality factors and personal values, *Personality and Social Psychology Bulletin*, 2002, 28(6), 789–801, DOI 10.1177/0146167202289008
- [43] A. M. Evans, W. Revelle, Survey and behavioral measurements of interpersonal trust, *Journal of Research in Personality*, 2008, 42(6), 1585–1593
- [44] K. M. DeNeve, H. Cooper, The happy personality: A meta-analysis of 137 personality traits and subjective wellbeing, *Psychological Bulletin*, 1998, 124(2), 197–229
- [45] K. Dautenhahn, Roles and functions of robots in human society: Implications from research in autism therapy, *Robotica*, 2003, 21(4), 443–452
- [46] K. Dautenhahn, S. Woods, C. Kaouri, M. L. Walters, K. L. Koay, I. Werry, What is a robot companion - friend, assistant or butler?, In: *Intelligent Robots and Systems Proceedings (IROS 2005)*, Edmonton, Alta., Canada, 2005, 1488–1493, DOI 10.1109/IROS.2005.1545189
- [47] S. Ljungblad, J. Kotrbova, M. Jacobsson, H. Cramer, K. Niechwiadowicz, Hospital robot at work: Something alien or an intelligent colleague? In: *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)*, New York, NY, USA, ACM, 2012, 177–186

- [48] M. Madsen, S. Gregor, Measuring human-computer trust, In: Proceedings of the 11th Australasian Conference on Information Systems, Brisbane, Australia, 2000, 53, 6–8
- [49] D. Kuchenbrandt, F. Eyszel, The mental simulation of a human-robot interaction: Positive effects on attitudes and anxiety toward robots, In: 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication, Paris, France, 2012, 463–468, DOI 10.1109/RO-MAN.2012.6343795
- [50] R. Wullenkord, F. Eyszel, Improving attitudes towards social robots using imagined contact, In: 23rd IEEE International Symposium on Robot and Human Interactive Communication, Edinburgh, UK, 2014, 489–494, DOI 10.1109/RO-MAN.2014.6926300