2020 17th International Conference on Ubiquitous Robots (UR)
June 22-26, 2020. Kyoto, Japan

Robot Agreeableness and User Engagement in Verbal Human-Robot Interaction

Luca Garello*,1, Francesco Grella*,1, Stefano Castagnetta*,1, Barbara Bruno[2],
Carmine Tommaso Recchiuto[1] and Antonio Sgorbissa[1]

***Abstract*— It is a well-established fact in the field of Human-Robot Interaction that the personality a social robot is endowed with (i.e., the one it expresses by its gestures and words) plays a key role towards the engagement experienced by any person interacting with it. In this article, we address the problem of designing the conversational style of a social robot to convey different degrees of *Agreeableness*, which is one of the Big Five personality traits, and investigate the relation between the person's engagement and the robot's degree of agreeableness.**

We propose to use subjective perception of time elapsed as an indicator for engagement and design two robot behaviours, one making the robot always agree with the person's opinion on topics of discussion (the *compliant* behaviour) and one making the robot always disagree (the *contrasting* behaviour). In an experiment involving 14 participants, we assess whether the robot adopting the *contrasting* behaviour is perceived as more engaging than the one adopting the *compliant* behaviour, and use participants' previous interactions with vocal assistants as a reference for estimating the naturalness, enjoyability and politeness of the conversation with the robot. Results suggest that designing social robots that are able to disagree with the person they're interacting with might be key to make them more engaging and entertaining.

I. Introduction

Engagement, defined as the process by which two or more participants establish, maintain and end their perceived connection [1], has long been established as a key factor for the success of any form of social interaction between a robot and a person, and especially of verbal interaction [1], [2].

Commonly adopted metrics for measuring engagement in Human-Robot Interaction (HRI) scenarios, derived from observations of Human-Human Interactions [3], include *gaze*-related metrics (how many, how long and how frequent are occurrences of directed gaze, shared gaze and mutual facial gaze), *speech*-related metrics (intended as the number of adjacency pairs, such as question-answer pairs, occurring during verbal conversations) and *gesture*-related metrics (typically intended as the number, type and frequency of backchannels, i.e., brief signals used to communicate understanding, confusion, desire for the other to continue...).

While the above indicators are sure to be highly relevant for the estimation of engagement, they are very difficult to reliably grasp via automated techniques. To overcome this obstacle, different strategies have been proposed in the literature. A number of field studies of engagement in HRI, for example, approximate gaze direction with head orientation and adjacency pairs with utterances and response time [4], [5], which are easier to detect than the original indicators but less clearly indicative of engagement. Following another approach, in an experiment with a robotic bartender engagement is estimated by means of features related to the pose of customers at the bar counter [6], under the intuition that domain-specific cues can be used as substitutes for the universal indicators. The main disadvantage of this approach is the need to identify and estimate reliable and detectable domain-specific cues, which also makes it harder to compare / generalise any obtained result onto different domains.

In this work, we propose to use *time*, and specifically the subjective perception of interaction time, as a reliable, general and easy-to-measure indicator for engagement in the context of HRI experiments. Subjective perception of time has been shown to be an effective indicator of cognitive involvement [7] and immersion [8], and there is evidence that "the more engaged someone is, the more likely they are to underestimate the passage of time" [9], [10].

To the best of our knowledge, this is the first time that subjective perception of time is considered as an indicator for engagement in the context of HRI.

Socially Assistive Robotics (SAR) is the research field aiming at the design of assistive robots (i.e., robots providing assistance to a user) which rely also, primarily, or even exclusively, on social and non-physical interaction to communicate with people [11]. Intuitively, the level of engagement of a person interacting with a social robot directly influences the quality, length and frequency of the interactions, and therefore the effectiveness of the robot in its assistive task.

As an example, the CARESSES project[1] is aimed at developing a socially assistive robots for the elderly, that relies on dialogue and culture-specific a-priori known information to drive its discovery of the user's habits and preferences [12]. In this case, ensuring that the user is engaged by the dialogue is crucial for the robot to discover and tailor its services and behaviour onto that person's preferences, which in turn is crucial for the effectiveness of the robot as an aide.

The identification of what, and to what extent, contributes to the engagement of the interaction between a person and a robot is an open and active research field.

Researchers have reached a broad consensus on the importance of *personality*, with recent studies specifically focusing

*The authors have contributed equally.
[1]DIBRIS, University of Genova, Italy.
[2]CHILI Lab & BioRob Lab, Swiss Federal Institute of Technology in Lausanne (EPFL), Switzerland. B. Bruno was with DIBRIS, University of Genova at the time this research was conducted.
Corresponding author's email: antonio.sgorbissa@unige.it

[1]http://caressesrobot.org/en/

978-1-7281-5715-3/20/$31.00 ©2020 IEEE

on the relationship between user personality, robot personality, and resulting engagement [13], [4]. Literature provides ample examples that (i) the person's personality plays a non-negligible role on their interaction with, and appreciation of, a robot [14] and (ii) the robot's personality has a similar effect, both when the personality is expressed via physical cues [15] and when it is expressed via verbal/behavioural cues [16]. Moreover, a growing number of research works investigates the relationship between user personality and robot personality [13], [17], specifically exploring whether similar personalities (e.g., leaning towards the same ends along the Big Five personality traits [18]) allow for a more effective and/or pleasurable interaction than complementary personalities (i.e., leaning towards opposite ends along the Big Five personality traits), or viceversa [2], [19], [16].

As far as verbal interaction is concerned, a large corpus of Literature focuses on the personality trait of *Extraversion* [18], specifically addressing how to design the robot's communication style to convey extraversion or introversion, and investigating the impact of this choice on the quality of the interaction. In a series of studies involving US, Chinese and German participants, people were asked to take decisions on an unfamiliar topic with the possibility of relying on the suggestions of a robot assistant [20], [21], [22]. Experimenters found that the communication style adopted by the robot (extraverted, i.e., explicitly providing its recommendation vs. introvert, i.e., implicitly hinting at it) has a direct impact on the number of times a person follows its advice, suggesting that people tend to trust more a robot which follows their same communication conventions.

In this article, we discuss a preliminary experimental framework to assess the impact of the personality trait of *Agreeableness*, proposing a design method for the robot's communication style to convey a friendly or challenging disposition, and investigating the impact of this choice on the quality of the interaction, specifically in terms of subjective perception of time, naturalness, enjoyability and politeness.

Concretely, we rely on the dialogue framework developed in the CARESSES project [23] to design two alternative verbal communication behaviours: a *compliant* behaviour, leading the robot to agree with the user's opinion on the topic of discussion, and a *contrasting* behaviour, leading the robot to disagree with the user's opinion. We adopt a between-subjects experimental design to compare the two behaviours and report the results obtained with 14 participants.

The article is structured as follows.

Section II explains the rationale behind the study, clarifying the research questions which drove the formulation of our hypotheses. In Section III the full description of the method is provided, while the full experimental setup description, comprehensive of design considerations and evaluation metrics, is given in Section IV. Sections V and VI respectively report and discuss the results obtained during the experiments and the insights they provided to drive future steps along this research direction. Conclusions follow.

II. Study Rationale

The research question inspiring this work is:

> *What is the relation between a social robot's agreeableness and the user's engagement?*

Concretely, addressing this question requires to: (i) identify a suitable metric for measuring the engagement; (ii) define, implement and test a rationale for designing a robot's communication style to convey different degrees of agreeableness; (iii) design an experiment allowing for measuring the influence, or lack thereof, of a robot's agreeableness on the engagement perceived by the person interacting with it.

Concerning the first issue, as discussed in the Introduction, we propose to rely on the person' subjective perception of time. Concerning the second issue, as detailed in Section III, we propose to consider the number of times the robot agrees with the user's opinion on a topic as the manipulable construct associated with agreeableness and define a *compliant* (very agreeable) behaviour as the one that always agrees with the user's opinion and a *contrasting* (very non agreeable) behaviour as the one that never agrees with the user's opinion on the given topic of discussion.

As discussed in the Introduction, literature reports of the correlation between challenges and engagement [7], which suggests that the *contrasting* robot might be perceived as more engaging than the *compliant* one. However, considering that engagement alone does not allow for distinguishing between a pleasurable and a non-pleasurable experience, we integrate the analysis of the subjective perception of time with that of the naturalness and enjoyability of the interaction, and introduce vocal assistants (such as Siri and Google Assistant) as a reference for participants to rate the quality of the interaction with the robot.

To perform an action or voice an opinion which is unexpectedly in contrast with others' social conventions and expectations is a well-known comedy mechanism to create entertaining characters and situations [24], that has been recently applied to create funny fictional robots as well. Marvin, the paranoid android of the novel "The Hitchhiker's Guide To The Galaxy"[2] is afflicted with severe depression (something we would not expect to see in a robot) and bleakly comments on any decision made by his more adventurous companions with sentences such as "I have a million ideas. But, they all point to certain death". Similarly, Bender, the robot from the animated sitcom Futurama by Matt Groening[3], is characterised by a bias towards non-robots, often underlined with the catchphrase "kill all humans".

The above research question is decomposed in our experiments into the following research hypotheses:

- **H1** Robots displaying a *contrasting* behaviour during verbal interactions are more engaging than robots displaying a *compliant* behaviour.
- **H2** Interactions with robots displaying a *contrasting* behaviour are perceived as more enjoyable than those with

[2] https://en.wikipedia.org/wiki/Marvin_the_Paranoid_Android

[3] www.youtube.com/watch?v=G1zv23yiDHc

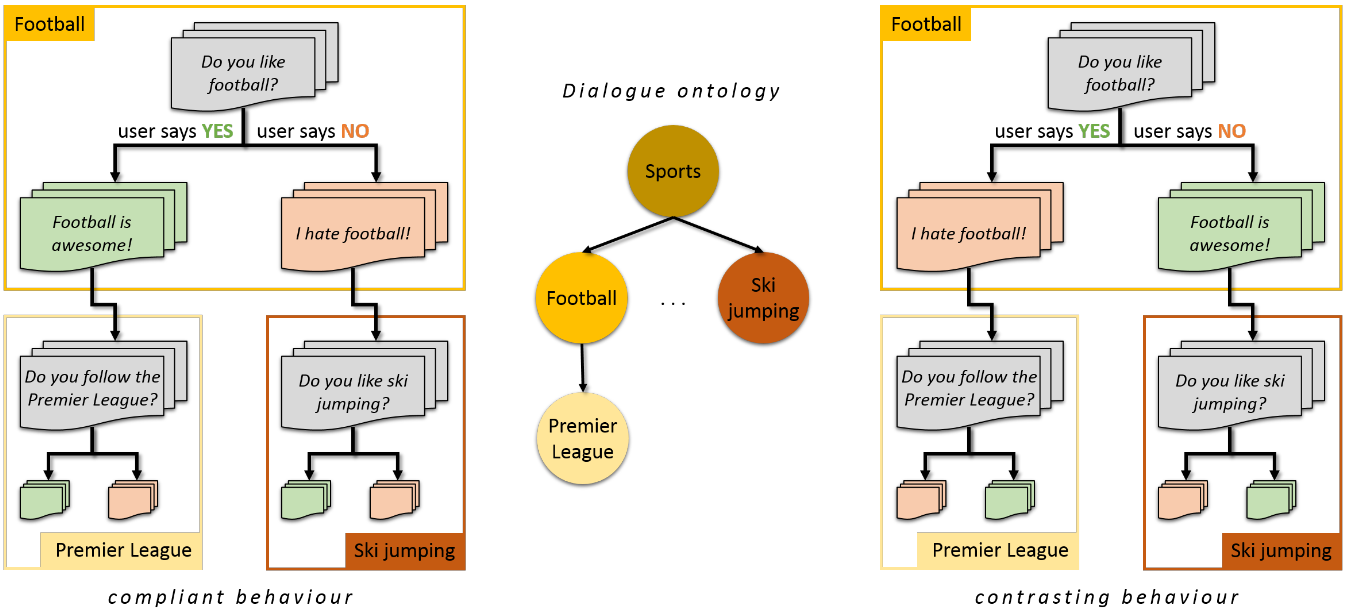

Fig. 1: Left: the dialogue pattern followed by the robot with *compliant* behaviour. Centre: a portion of the structure encoding available topics of discussion and related information. Right: the dialogue pattern followed with the *contrasting* behaviour.

robots displaying a *compliant* behaviour.

H3 Verbal interactions with robots, regardless of the communication style, are perceived as more natural and more enjoyable than those with vocal assistants.

The experimental framework designed to evaluate the above hypotheses is detailed in Section IV.

III. Method

In this work, we adopt the dialogue framework developed within the CARESSES project to endow a robot with the ability to autonomously discuss a wide variety of topics of conversation with a person [23], tailoring it to suit communication styles with different degrees of agreeableness.

The framework relies on an ontology, which is a formal naming and definition of the types, properties, and interrelationships of the entities relevant for a particular domain of discourse [25]. The terminology defining the domain of discourse, containing general properties of concepts, is stored in the terminological box (TBox) of the ontology, while knowledge that is specific to instances belonging to the domain is stored in the assertional box (ABox).

In our case the Tbox contains concepts describing possible topics of conversations (such as those shown in the centre of Figure 1), and a special property (hasTopic) links them one another in a hierarchical tree. Concretely, the purpose of the hasTopic property and of the tree structure is to ensure that concepts are brought up by the robot in a conversation in “plausible” sequences for the human counterpart.

In the ABox, an instance is created for each concept to be discussed, and a number of dialogue-related properties are filled. Specifically:

- hasQuestion contains the question(s) the robot can use to ask the user about the instance. Questions are denoted with grey boxes in Figure 1. As an example, one of the questions associated with the concept *Football* is “Do you like football?”;
- hasPositiveSentence contains the sentence(s) the robot can use to express a positive attitude towards the instance. Positive statements are denoted with green boxes in Figure 1. As an example, one of the positive statements associated with the concept *Football* is “Football is awesome!”;
- hasNegativeSentence contains the sentence(s) the robot can use to express a negative attitude towards the instance. Negative statements are denoted with red boxes in Figure 1. As an example, one of the negative statements associated with the concept *Football* is “I hate football!”.

The diagrams at the sides of Figure 1 sketch the dialogue patterns followed by the robot, in the case it is adopting the *compliant* behaviour (on the left side of the Figure) or the *contrasting* behaviour (on the right side of the Figure).

In both cases, once the robot has identified the topic to discuss with the user, it introduces it by asking one randomly chosen question among those encoded in the hasQuestion property and then parses the person’s response looking for keywords denoting a positive (“YES” arrows in the Figure) or negative attitude towards the topic (“NO” arrows in the Figure). Then, the robot’s response differs according to the adopted behaviour:

- the *compliant* behaviour leads the robot to always agree with the user by responding with a randomly chosen statement matching the user’s attitude (i.e., a sentence among those encoded in the hasPositiveSentence property if the user expressed a positive attitude towards the topic, or one among those encoded in the hasNegativeSentence property if the user expressed a negative attitude);

- the *contrasting* behaviour leads the robot to always disagree with the user by responding with a randomly chosen statement opposing the user's attitude (i.e., a sentence among those encoded in the hasNegativeSentence property if the user expressed a positive attitude towards the topic, or one among those encoded in the hasPositiveSentence property if the user expressed a negative attitude).

In both cases, after uttering its response, the robot again considers the user's opinion to decide whether to pick a concept that follows the previous one in the hierarchical tree (such as *Premier League* in the example of Figure 1) or jump to another, unrelated topic (such as *Ski jumping* in the example of Figure 1). Once the concept to discuss is chosen, the robot randomly selects the associated question to ask and repeats the above-outlined procedure.

The selection of the sentences to be encoded in the hasPositiveSentence and hasNegativeSentence properties of concepts deserves a special attention. To mitigate the effects of the *Extraversion* personality trait (on the robot' side as well as on the user' side), which, as discussed in the Introduction, influences the strength with which opinions are conveyed, we have identified four possible categories of robot responses, ranging from very explicit and personal to implicit and impersonal:

- a *strong* sentence expresses the robot's personal opinion without explaining it;
- an *experience*-related sentence provides personal evidence supporting the robot's opinion;
- a *joke* conveys the robot's opinion, possibly implicitly, with a humorous comment;
- an *advice* provides factual, impersonal evidence supporting the robot's opinion, that is not explicitly stated.

Figure 2 provides some examples for each type of sentence, ordered from one displaying a very explicit positive attitude (top line in the Figure) to one displaying a very explicit negative attitude (bottom line in the Figure).

For all concepts considered in our experiment, the properties hasPositiveSentence and hasNegativeSentence are filled with multiple sentences, belonging to different types. During interaction, as outlined above, the behaviour determines the group to pick a sentence from, while the specific sentence uttered by the robot is chosen randomly. Specifically, the *contrasting* robot can pick sentences of all types, while we have constrained the *compliant* robot to only use mild statements. As a result, in both cases some of the robot's responses are more extravert than others, and some are closer to the person's degree of extraversion than others, thus mitigating the effect of the robot's and the user's extraversion on the interaction.

The list of topics to discuss, and their order, is fixed, to ensure that the robot's dialogue pattern is the same between the two experimental arms. Specifically, topics of conversation belong to the categories of *Pets, Sports, Music,*

Fig. 2: Examples of sentences encoded in the hasPositiveSentence property of topics of conversations (top half), and sentences encoded in the hasNegativeSentence property of topics of conversations (bottom half). Sentences at the extremes convey the robot's opinion more explicitly.

Fig. 3: Experimental setup. Once the interaction is over, the participant is asked to move to the desk on the right and fill in an online questionnaire. Researchers, not shown, stand close by to intervene upon request.

Religion, Health, Clothes, Personal care and are addressed in the conversation with one general question and one optional more specific question if the person agrees to proceed along that topic. Each concept is associated with at least 3 sentences for the hasPositiveSentence property and 3 for the hasNegativeSentence property, belonging to different types among those discussed above. All sentences have been chosen and manually encoded by the researchers, ensuring that they all have comparable length.

IV. Experimental Setup

A. Study variables

The independent variable of the study is the conversation behaviour of the robot, which can either be *compliant* or *contrasting*, and the purpose of the experiment is to evaluate its impact on the subjective perception of time, which is the main dependent variable, along with naturalness and enjoyability of the conversation. We identify the robot's extraversion, the chosen topics of conversation, familiarity with robots, familiarity with vocal assistants and age as possible confounders. The former two variables pertain to the design of the robot's dialogue management algorithm and the method proposed to mitigate their impact on the interaction is discussed in Section III. The latter variables pertain to the participants' background and their impact has been mitigated by assigning them to the two arms equally. The results of the manipulation check on these variables are reported in Section V-A.

B. Study Design

The main research hypothesis, [H1], outlined in Section II, is evaluated with a between-group study with two experimental arms, one in which participants interact with a robot adopting the *compliant* behaviour and the other in which participants interact with the same robot, adopting the *contrasting* behaviour.

Figure 3 shows the area in which the experiment takes place. Upon arrival, the participant is welcomed by a researcher and instructed on what to do. Specifically, all participants are told that the experiment consists in chatting with the robot (Pepper, produced by Softbank Robotics Europe) and it is devoted to assessing the robot's conversational capabilities.

According to the arm to which the participant is assigned, the robot is set up to use the *compliant* or *contrasting* behaviour during the interaction, and each interaction is timed to last 10 minutes. Researchers stand close by for the whole duration of the interaction, to intervene upon request.

In both arms, at the end of the interaction, participants are asked to move to a PC desk and fill in a questionnaire[4] designed ad-hoc for this study, which includes items related to subjective perception of time elapsed [H1] and naturalness and enjoyability of the conversation [H2]. Items related to the latter two categories appear twice in the questionnaire, once referred to the interaction with the robot and once referred to previous interactions with vocal assistants [H3].

C. Measures and Metrics

The questionnaire[4] is organised in three sections. The first one includes background items, the second one refers to previous interactions with vocal assistants and includes the 8 items listed in Table I, presented as 10-points Likert scales, while the third section refers to the interaction with the robot and includes the same 8 items listed in Table I, again presented as 10-points Likert scales. The third section includes an additional item in which the participant is asked to estimate the duration of the interaction with the robot. Questionnaire items address the naturalness of the conversation (items 1,3,7), the enjoyability of the conversation (items 2,5,6) and the politeness of the robot's behaviour (items 4,8).

[4]https://forms.gle/xWtDcYTHJMo2qBZr8

TABLE I: Questionnaire Items

#	Description
(1)	Feeling of being understood
(2)	Level of interest of the conversation
(3)	Level of coherency of the answers
(4)	How comfortable was the conversation?
(5)	Level of emotional involvement
(6)	Level of enjoyment of the conversation
(7)	How natural was the conversation?
(8)	Level of courtesy

V. Results

A. Population

A total of 14 individuals (9 females and 5 males), aged 20 to 29 years old, participated in the experiment, equally split in the two arms. With the sole exception of one participant they all use vocal assistant (e.g., Google Assistant, Alexa, Siri...), most of them with the primary purpose of getting information. All but one participant are university students. The group interacting with the *contrasting* robot includes 3 people with familiarity with robots (and 4 who never interacted with one), while the group interacting with the *compliant* robot includes 4 people with familiarity with robots (and 3 who never interacted with one).

Concerning the mitigation measures for possible confounders related to the participants' background, while the two arms are very similar for the participants' age, gender and familiarity with robots, there is a slight difference in terms of familiarity with vocal assistants. Specifically, while all participants assigned to the *contrasting robot* group have previous experiences with vocal assistants, one participant assigned to the *compliant robot* group had never interacted with a vocal assistant. The responses of this participant (S-7) have been excluded from the analysis reported in the following Sections. A number of consideration arising from this case are discussed in Section VI-C.

B. Engagement - Subjective perception of time

The average subjective perception of time elapsed for the participants interacting with the *compliant* robot is 10.3 minutes (std = 2.58 minutes), while the average subjective perception of time elapsed for the participants interacting with the *contrasting* robot is 9.0 minutes (std = $3,27$ minutes). A two-tailed t-test indicated no statistically-significant effect of the robot behaviour on the subjective perception of time elapsed [H1]. Please notice that the response given by the participant who never interacted with a vocal assistant (S-7) is not included in the control group average.

C. Naturalness, enjoyability and politeness

We performed a Kruskal-Wallis H test, useful when dealing with non-gaussian distributions, to check for statistically significant differences in the questionnaire items responses: the results are reported in Tables II and III.

The leftmost section of Table II ("cp / cr" columns) and Figure 4 investigate whether the naturalness and enjoyability

TABLE II: Analysis of the naturalness, enjoyability and politeness scores given to the *compliant* robot (cp) w.r.t. the *contrasting* robot (cr), and to robots (r) w.r.t. vocal assistants (va). Bolded items represent statistically significant differences ($\alpha = .05$, $p < .05$).

	cp / cr		r / va	
	χ^2	p	χ^2	p
(1)	0.02	0.8936	0.17	0.6767
(2)	0.64	0.4227	3.04	0.0815
(3)	1.01	0.3157	**7.52**	**0.0061**
(4)	1.10	0.2936	0.89	0.3452
(5)	0.07	0.7890	**7.63**	**0.0057**
(6)	0.42	0.5169	**6.60**	**0.0102**
(7)	0.11	0.7362	**14.32**	**0.0002**
(8)	0.53	0.4661	2.02	0.1548

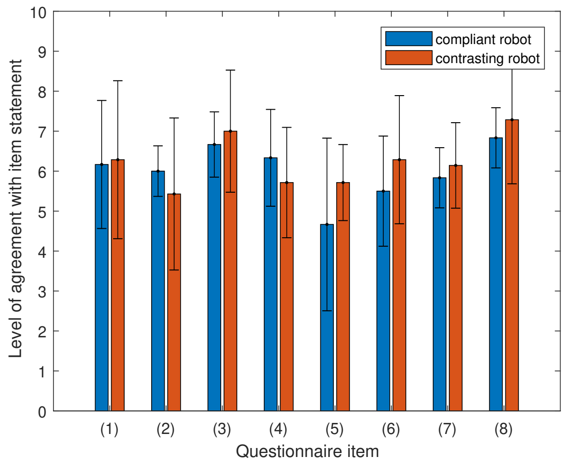

Fig. 4: Participants' assessment of the interaction with the robot: comparison between the *compliant* (blue) and *contrasting* (red) behaviours. Question items refer to those listed in Table I.

of conversations with the two robots are similar or not [H2], by comparing the questionnaire responses concerning the interaction with the robot given by participants interacting with the *compliant* robot (blue columns in the Figure) with those given by participants interacting with the *contrasting* robot (red columns in the Figure). As the results show, there are no statistically significant differences between the two robots, and both are rated very highly.

The rightmost section of Table II ("r / va" columns) and Figure 5 investigate whether the naturalness and enjoyability of conversations with the robots are similar or not to those of conversations with vocal assistants [H3], by comparing the questionnaire responses given by all participants to the interaction with the robot (red columns in the Figure) with those they gave to previous interactions with vocal assistants (blue columns in the Figure). As the results show, there are statistically significant differences in the scores of items (3), (5), (6), (7), with previous interactions with vocal assistants being evaluated as significantly worse than the interaction with the robot, especially in naturalness and enjoyability.

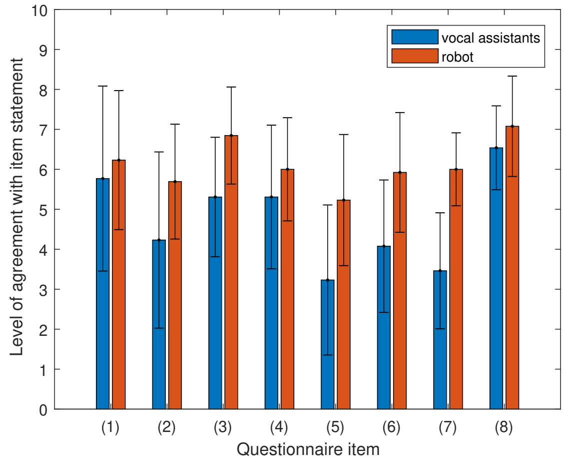

Fig. 5: Participants' assessment of the interaction with the robot w.r.t. previous interactions with vocal assistants. Question items refer to those listed in Table I.

TABLE III: Analysis of the naturalness, enjoyability and politeness scores given to the *compliant* robot (cp) w.r.t. the vocal assistants (va), and to the *contrasting* robot (cr) w.r.t. vocal assistants (va). Bolded items represent statistically significant differences ($\alpha = .05$, $p < .05$).

	cp / va		cr / va	
	χ^2	p	χ^2	p
(1)	0.33	0.5648	1.23	0.2683
(2)	0.01	0.9318	**3.89**	**0.0487**
(3)	1.88	0.1698	**5.37**	**0.0205**
(4)	0.33	0.5683	0.52	0.4713
(5)	0.97	0.3245	**7.82**	**0.0052**
(6)	0.68	0.4108	**6.72**	**0.0095**
(7)	**4.32**	**0.0376**	**9.08**	**0.0026**
(8)	0.03	0.8629	2.52	0.1126

Differences in politeness do not appear to be significant.

As a follow-up to the above analysis, Table III and Figure 6 investigate whether there are differences in the comparison between the interaction with the robot and previous interactions with vocal assistants between participants belonging to the *compliant* group (blue columns in the Figure) and those belonging to the *contrasting* group (red columns in the Figure). As the Table and Figure show, participants who interacted with the *compliant* robot only report previous interactions with vocal assistants to be significantly less natural than the conversation with the robot (item 7), while participants who interacted with the *contrasting* robot report previous interactions with vocal assistants to be not only less natural (items 3 and 7), but also significantly less enjoyable (items 2, 5 and 6) than the conversation with the robot.

VI. Discussion

A. Engagement - Subjective perception of time

The preliminary results obtained in the study do not conclusively support hypothesis [H1] that robots displaying a *contrasting* behaviour during verbal interactions are more

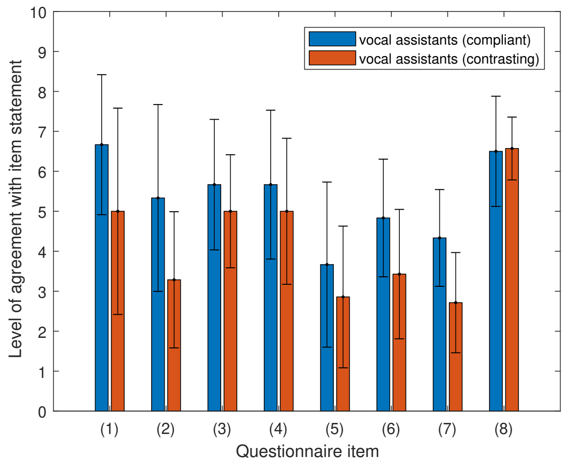

Fig. 6: Participants' assessment of previous interactions with vocal assistants: comparison between the responses of participants in the *compliant* group (blue) and those of participants in the *contrasting* group (red). Question numbers refer to the items listed in Table I.

engaging than robots displaying a *compliant* behaviour, although the observed difference of approximately 10% between the two arms in favour of the former encourages further experiments with a broader population.

We identify the duration of the interaction with each participant, which is fairly short and a round number (10 minutes) to be a factor that could have played a levelling-out effect and plan, for future studies, to extend it to a number above 15 minutes that is not a multiple of 5.

B. *Naturalness, enjoyability and politeness*

The results obtained in the study provide preliminary support to the hypothesis [H3] that verbal interactions with robots are perceived as more natural and enjoyable than those with vocal assistants, as shown in Table II and Figure 5.

It is very interesting to notice that there is no significant difference in the assessment of the interaction with the robot between the two arms (as it appears from the "cp / cr" columns of Table II and Figure 4), although the two robots behave very differently. Participants don't seem to find incoherent or unnatural the conversation with the *contrasting* robot (items 3 and 7), which suggests that its dialogue pattern is not perceived as weird. Moreover, results suggest that the *contrasting* robot is not perceived as less polite than the *compliant* one (see Figure 4), and both are rated as polite as vocal assistants (see Figure 5).

Conversely, there is a significant difference between the two arms in the assessment of previous interactions with vocal assistants, with participants interacting with the *contrasting* robot judging previous conversations with vocal assistants remarkably less enjoyable than participants interacting with the *compliant* robot. In our study, participants first interacted with the robot, then were asked to rate previous interactions with vocal assistants, and lastly to rate the interaction with the robot. We hypothesise that participants used the naturalness and enjoyability of the conversation with the robot as a reference for rating the interactions with vocal assistants, and this fact led those assigned to the *compliant* robot to rate previous interactions with vocal assistants as "less enjoyable, although not much" than the one just sustained with the robot, and those assigned to the *contrasting* robot to rate previous interactions with vocal assistants as "remarkably less enjoyable" than the one just sustained with the robot.

If confirmed, this finding would provide support for our hypothesis [H2] and, more generally, prove that the degree of *Agreeableness* of a robot, as expressed via its communication style, has a huge impact on the quality of the interaction.

To proceed along this research line, we are planning additional studies, specifically: (i) with longer conversations, to mitigate the novelty effect, and (ii) better controlling for familiarity with the language spoken by the robot (English). Concerning this last issue, although none of the participants of our experiment were native English speaker, they all possess the same certification of knowledge of the English language. However, we observed that some participants asked the robot to repeat its statements more than others, and that sentences encoded in the hasPositiveSentence and hasNegativeSentence have different complexity (some include words more common than others). Standardizing the language and using simple words will help with all english speaking demographics for future studies.

C. *No previous interaction with vocal assistants*

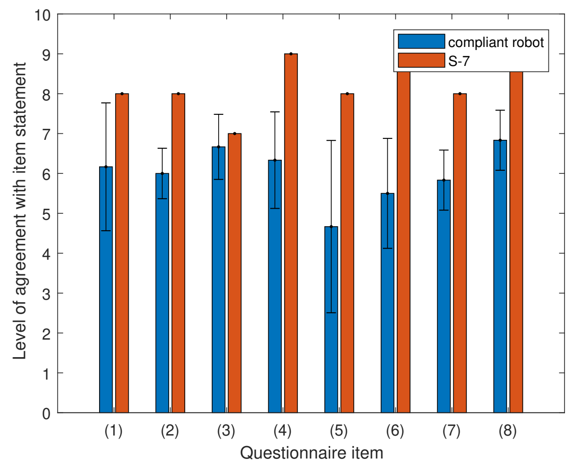

Fig. 7: Participants' assessment of the interaction with the robot: comparison between S-7 (red) and the other people interacting with the *compliant* robot (blue). Question numbers refer to the items listed in Table I.

As anticipated in Section V-A, participant S-7, assigned to the *compliant robot* group, reported to have never interacted with vocal assistants and thus did not complete the corresponding questionnaire section. His/her responses have been analysed separately from the rest of the group.

Concerning the subjective perception of time, the estimate of S-7 is of 3 minutes, which is significantly lower than all other participants' estimates and an interesting indication of the relevance of this metric. Figure 7 compares the assessment of the naturalness, enjoyability and politeness of the robot given by S-7 (red columns in the Figure) with the assessment given by the other participants in the *compliant robot* group (blue columns in the Figure), showing that S-7 rated the interaction with the robot remarkably more positively. This comparison can be useful to estimate the influence of the novelty effect on participants and suggests that familiarity with vocal assistants might help mitigate it.

VII. Conclusions

The article investigates the relation between the *Agreeableness* of a social robot and the *Engagement* perceived by a person interacting with it. We focus on verbal interaction exclusively and associate *Agreeableness* with the number of times the robot agrees with the person's opinion in the context of a conversation. We rely on the dialogue framework developed in the CARESSES project to design a behaviour leading the robot to always agree with the user (which we call *compliant*) and a behaviour leading the robot to always disagree with the user (which we call *contrasting*).

We have designed a study to evaluate whether the robot adopting the *contrasting* behaviour is perceived as more engaging than the one adopting the *compliant* behaviour, using the participants' subjective perception of time elapsed as primary indicator of engagement, and preliminarily evaluate the naturalness, enjoyability and politeness of the two robots using conversations with vocal assistants as a reference.

Experiments with 14 participants suggest that the research direction is well worth exploring and that designing social robots that are able to disagree with the person they're interacting with might be a key feature to make them more engaging and entertaining.

VIII. Acknowledgement

This work has been partially supported by the European Commission Horizon2020 Research and Innovation Programme under grant agreement No. 737858.

References

[1] C. Sidner and M. Dzikovska, "Human-robot interaction: engagement between humans and robots for hosting activities," in *Proceedings. Fourth IEEE Int. Conf. on Multimodal Interfaces*, no. May. IEEE Computer Society, 2002, pp. 123–128.

[2] A. Aly and A. Tapus, "Towards an intelligent system for generating an adapted verbal and nonverbal combined behavior in human–robot interaction," *Autonomous Robots*, vol. 40, no. 2, pp. 193–209, 2016.

[3] C. Rich, B. Ponsler, A. Holroyd, and C. L. Sidner, "Recognizing engagement in human-robot interaction," in *Proceeding of the 5th ACM/IEEE Int. Conf. on Human-robot interaction - HRI '10*. New York, New York, USA: ACM Press, 2010, pp. 375–382.

[4] S. Ivaldi, S. Lefort, J. Peters, M. Chetouani, J. Provasi, and E. Zibetti, "Towards Engagement Models that Consider Individual Factors in HRI: On the Relation of Extroversion and Negative Attitude Towards Robots to Gaze and Speech During a Human–Robot Assembly Task," *International Journal of Social Robotics*, vol. 9, no. 1, pp. 63–86, 2017.

[5] S. M. Anzalone, S. Boucenna, S. Ivaldi, and M. Chetouani, "Evaluating the Engagement with Social Robots," *International Journal of Social Robotics*, vol. 7, no. 4, pp. 465–478, 2015.

[6] M. E. Foster, A. Gaschler, and M. Giuliani, "Automatically Classifying User Engagement for Dynamic Multi-party Human–Robot Interaction," *International Journal of Social Robotics*, vol. 9, no. 5, pp. 659–674, 2017.

[7] D. Baldauf, E. Burgard, and M. Wittmann, "Time perception as a workload measure in simulated car driving," *Applied Ergonomics*, vol. 40, no. 5, pp. 929–935, 2009.

[8] C. Jennett, A. L. Cox, P. Cairns, S. Dhoparee, A. Epps, T. Tijs, and A. Walton, "Measuring and defining the experience of immersion in games," *International Journal of Human-Computer Studies*, vol. 66, no. 9, pp. 641–661, 2008.

[9] S. Attfield, G. Kazai, M. Lalmas, and B. Piwowarski, "Towards a science of user engagement (Position Paper)," in *WSDM'11*. ACM Press, 2011.

[10] A. M. Sackett, T. Meyvis, L. D. Nelson, B. A. Converse, and A. L. Sackett, "You're Having Fun When Time Flies: The Hedonic Consequences of Subjective Time Progression," *Psychological Science*, vol. 21, no. 1, pp. 111–117, 2010.

[11] D. Feil-Seifer and M. Mataric, "Socially assistive robotics," *IEEE Robotics and Automation Magazine*, vol. 18, no. 1, pp. 24–31, 2011.

[12] B. Bruno, N. Y. Chong, H. Kamide, S. Kanoria, J. Lee, Y. Lim, A. K. Pandey, C. Papadopoulos, I. Papadopoulos, F. Pecora, A. Saffiotti, and A. Sgorbissa, "Paving the way for culturally competent robots: A position paper," in *RO-MAN 2017 - 26th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2017, pp. 553–560.

[13] H. Salam, O. Celiktutan, I. Hupont, H. Gunes, and M. Chetouani, "Fully Automatic Analysis of Engagement and Its Relationship to Personality in Human-Robot Interactions," *IEEE Access*, vol. 5, pp. 705–721, 2017.

[14] A. Cruz-Maya and A. Tapus, "Teaching nutrition and healthy eating by using multimedia with a Kompai robot: Effects of stress and user's personality," in *2016 IEEE-RAS 16th Int. Conf. on Humanoid Robots (Humanoids)*. IEEE, nov 2016, pp. 644–649.

[15] J. Hwang, T. Park, and W. Hwang, "The effects of overall robot shape on the emotions invoked in users and the perceived personalities of robot," *Applied ergonomics*, vol. 44, no. 3, pp. 459–471, 2013.

[16] K. M. Lee, W. Peng, S.-A. Jin, and C. Yan, "Can Robots Manifest Personality?: An Empirical Test of Personality Recognition, Social Responses, and Social Presence in Human–Robot Interaction," *Journal of Communication*, vol. 56, no. 4, pp. 754–772, 2006.

[17] L. P. Robert Jr., "Personality in the Human Robot Interaction Literature: A Review and Brief Critique," in *Twenty-fourth Americas Conference on Information Systems*, 2018.

[18] L. R. Goldberg, "An alternative "description of personality": The Big-Five factor structure." *Journal of Personality and Social Psychology*, vol. 59, no. 6, pp. 1216–1229, 1990.

[19] O. Celiktutan and H. Gunes, "Computational analysis of human-robot interactions through first-person vision: Personality and interaction experience," in *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, aug 2015, pp. 815–820.

[20] V. Evers, H. Maldonado, T. Brodecki, and P. Hinds, "Relational vs. group self-construal: untangling the role of national culture in HRI," in *Human-Robot Interaction (HRI), 2008 3rd ACM/IEEE Int. Conf. on*. IEEE, 2008, pp. 255–262.

[21] P. P. Rau, Y. Li, and D. Li, "Effects of communication style and culture on ability to accept recommendations from robots," *Computers in Human Behavior*, vol. 25, no. 2, pp. 587–595, 2009.

[22] L. Wang, P.-L. P. Rau, V. Evers, B. K. Robinson, and P. Hinds, "When in rome: the role of culture & context in adherence to robot recommendations," in *Proceedings of the 5th ACM/IEEE Int. Conf. on Human-robot interaction*. IEEE Press, 2010, pp. 359–366.

[23] B. Bruno, C. T. Recchiuto, I. Papadopoulos, A. Saffiotti, C. Koulouglioti, R. Menicatti, F. Mastrogiovanni, R. Zaccaria, and A. Sgorbissa, "Knowledge Representation for Culturally Competent Personal Robots: Requirements, Design Principles, Implementation, and Assessment," *International Journal of Social Robotics*, 2019.

[24] D. Madden, *Harlequin's stick, Charlie's cane: a comparative study of commedia dell'arte and silent slapstick comedy*. Popular Press, 1975.

[25] N. Guarino, "Formal Ontology and Information Systems," in *Proceedings of FOIS'98*. Trento, Italy: IOS Press, 1998, pp. 3–15.