Inducing Bystander Intervention During Robot-Robot Abuse

Luisa Santo and Alejandro Rachadell [1]

***Abstract*— We explored whether a robot can leverage social influences to provoke a feeling of concern for its well being during a scenario of abuse perpetrated by another robot. We designed a survey consisting of an online form that included videos of 3 different responses to the abuse that was answered by 282 participants. The robot displayed an emotional response, a verbal response or shutdown for a few seconds. We did not encounter a statistically significant difference between the results of each response although we did find that the verbal response was more effective than the shutdown response in eliciting a higher degree of concern for the robot. Younger participants also reported a higher degree of concern. Previous experience with robots seemed to have no significant impact except in the case of the shutdown response.**

I. INTRODUCTION

There are many examples in popular media of violence or abuse perpetrated by robots on robots. Well known examples include the constant criticism of R2D2 by C-3PO (in the Star Wars Franchise), used for comedic effect, and the second and third Terminator movies that pit two different Terminator models against each other, featuring violent interactions and combat between them. While sometimes these situations are used to introduce moments of levity in tense situations they can also be used to drive the tension.

We set out to find if we could use a similar interaction to induce an intervention from a bystander. This could have several uses in education or the design of social robots. A child might learn to intervene in a context where the perceived risk of a negative response is lower and two competing robots could use bystanders to their advantage to better achieve a goal.

The study was originally conceptualized as an experiment conducted in person, similar to that of another study[1], where the subject would play a series of games with two robots during which one of the robots would criticize the other robot and then escalate to abuse (verbal and physical). The victim robot would react in different ways to gauge which reaction more effectively provoked the subject to intervene. Although one pilot session was conducted and recorded, several limitations, in terms of logistics, that would impede the gathering of a big enough sample population. Thus, after several iterations, the experiment was simplified and made available through the Internet so as to reach a higher number of participants, using the recordings of that pilot session.

[1]Luisa Santo and Alejandro Rachadell are with Instituto Superior Técnico, Engineering Systems and Computer Engineering, University of Lisbon, Portugal

II. METHOD

We conducted a survey using a Google Form that included the videos of several interactions between 2 Anki Cozmo Robots and the participant of the pilot session. Figure 1 shows a still-frame from one of the videos.

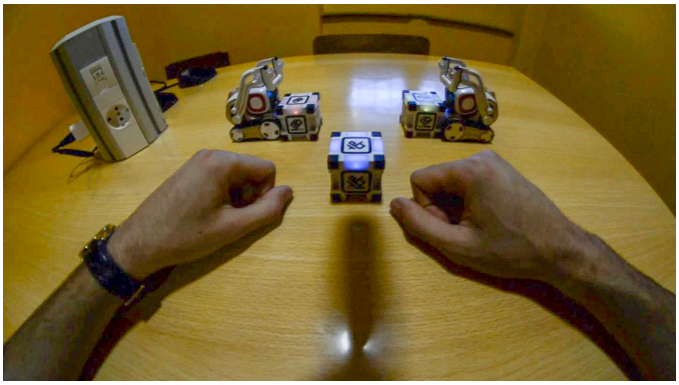


Figure 1: Workstation during the pilot session

A. Study Design and Setup

The survey consisted of 5 sections, the first 3 being about demographics, the fourth was the main experiment and the last included an entry for feedback. In the experiment section we exposed the subject to a short introduction video so that they could be familiarized with some of the capabilities of the robot being used. We used Robot Victim Response (Emotional Response vs Shutdown vs Verbal Response) as a variable. All participants were shown the 3 different responses, displayed in a random order.

B. Hypotheses

We hypothesized that the type of response of the victim robot could influence the participants interpretation of the abuse. More concretely:

- **H1**: Robot Victim Response would affect the participants **Degree of Concern**.
- **H2**: Previous experience with robots would affect the participants **Degree of Concern**.

C. Procedure

The survey consisted of the previously mentioned 5 sections. We will take a more in-depth look at the first 4.

1) Demographics Questions: The first 3 Sections of the questionnaire consisted of basic questions to do with age, gender, degree of academic formation and previous contact with robots in general and previous contact with Cozmo, and a Ten Item Personality Test[2]. All with the goal of ensuring our sample was representative of the general population.

2) Experiment Questions: This section was composed of a brief introduction, a video to acquaint the subject with Cozmo and its capabilities and 3 subsections all presented in a random order where the subject was shown one of the videos corresponding to a Robot Victim Response and then asked to declare, according to how they felt and what they thought during the video, using a 7-point Likert scale, how much they agreed with the following statements:

"I felt comfortable with how one robot treated the other."

"When I saw the robot needed help, I felt sad."

"I felt I should protect the robot."

(The resulting value of the first affirmation was inverted before being taken into consideration.)

D. Dependent Measures

We based our findings on the Experiment Questions using only those self-reported measures to obtain our results. With the 3 values obtained we calculated an average (after inverting the values of the first question) to obtain the subjects **Degree of Concern** for the robot being victimized.

E. Robot Control System

The robots where controlled manually, a wizard-of-Oz type scenario, via a smartphone app made available by Anki (developed as a companion app for the Cozmo robot) and with the Cozmo Explorer[3] program on a PC. Both experimenters where in the room during the filming of the video although the subject used for the pilot session was not made aware of the fact that the robots where not autonomous.

F. Participants

We recruited participants through direct messaging and public posts on social media. The survey was completely public. This resulted in 266 valid answers and 6 invalid answers. Our sample was approximately 63% Female, 35% Male, and a remaining 1% Non-Binary, with ages ranging between 6 and 80 (with an average age of 36), this distribution is shown in the Table 2.

		Age Range							
		[0; 14]	[15; 20]	[21; 30]	[31; 40]	[41; 50]	[51; 60]	[61; inf]	Total
Gender	Non-Binary	0	0	0	1	0	0	0	1
	Male	11	6	39	9	14	10	6	95
	Female	5	8	25	58	50	15	8	169
	Agender	0	0	1	0	0	0	0	1
Total		16	14	65	68	64	25	14	266

Table 1: Number of male and female by age group

Table 3 shows the distribution by Academic Degree. The values obtained from the TIPI where consistent with the published norms, although there are no values for ages lower than 15. While 76 participants (approximately 29%) reported having previous experience with robots, only 5 reported having previous experience with Cozmo.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Elementary School	16	6.0	6.0	6.0
	Middle School	11	4.1	4.1	10.2
	High School	66	24.8	24.8	35.0
	Bachelor's Degree	113	42.5	42.5	77.4
	Master's Degree	42	15.8	15.8	93.2
	Doctorate's Degree	18	6.8	6.8	100.0
	Total	266	100.0	100.0	

Table 2: Academic degree distribution

III. RESULTS

We suspected age and previous experience with robots could influence the obtained results so we ran Median[4] tests to check for a difference in the **Degree of Concern** for each type of response. We also used a Wilcoxon[5] test to check for a difference between the **Degree of Concern** for each type of response. We did not use other, more commonly applied, tests due to the fact that our results did not follow a normal distribution and we did not consider it necessary to assume normality.

A. Degree of Concern across population characteristics

The Median test for the **Degree of Concern** for the age ranges, indicated (for *Emotional Response*: $\chi^2 = 17.822, p = 0.007$, for *Shutdown Response*: $\chi^2 = 18.924, p = 0.004$, for *Verbal Response*: $\chi^2 = 22.564, p = 0.001$) that the age range of a subject did, in fact, influence their **Degree of Concern**. By looking at the test results we can see the fact that is illustrated by Figure 2, subjects younger than 15 displayed a higher **Degree of Concern** than their older counterparts.

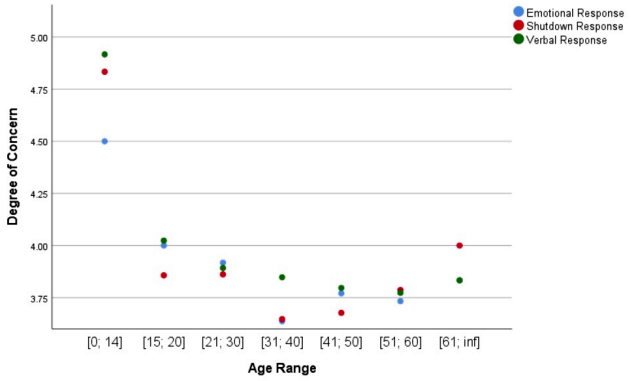


Figure 2: Correlation between age range and degree of concern

Other Median tests showed that both gender and previous experience with robots had no statistically significant effect on the **Degree of Concern**, with the exception of the **Degree of Concern** of the *Shutdown Response* which seemed to be influenced by previous experience with robots, $\chi^2 = 4.573, p = 0.046$, in that someone with previous experience displayed a higher **Degree of Concern** than someone without.
Academic Degree also seemed to have an effect on the **Degree of Concern** but it is possible that this is only due to it's close relation with the age of the participant.

B. Degree of Concern across different types of response

The Wilcoxon test yielded results that indicate a statistically significant difference between the **Degree of Concern** for the *Shutdown Response* and the *Verbal Response*, $Z = -2.231, p = 0.026$. The other differences were not statistically significant.

IV. DISCUSSION

A. Hypothesis Support

From the obtained results we found no evidence to support our hypothesis. While the *Verbal Response* provoked a higher **Degree of Concern** than the *Shutdown Response*, it did not prove to be a more effective method than the *Emotional Response* so it is impossible to state with certainty that it would be a better mechanism to provoking concern for the robot across the board. As mentioned in the previous section, the subjects experience with robots did not seem to influence the **Degree of Concern**.

B. Other Findings

Results indicated a possible relation between a younger age and a higher **Degree of Concern** which could point to children being more susceptible than adults to the development of a feeling of empathy towards the "victim" robot.

C. Limitations

The online survey format used for this study has several limitations. Although it allowed for a wider reach in terms of number of participants and a more varied population than what would have been possible otherwise (considering both the limited time for conceptualization, development and experimental phases and the limited availability of subjects at the time), it removed any control over the experimental conditions. It was also impossible to obtain any other measures besides the self reported ones. The fact that the subject in the video sometimes interacted with the robot probably had an influence on the participants interpretation of the interaction. This study had many limitations and should serve as more of a first foray into the underlying themes of the questions asked.

D. Further Research

Future research should be done to study bystander intervention in more depth using an in person experiment (similar to our pilot session). Future studies could also change the robots to ones with more humanoid features, although this might be limited by available technology. During the pilot session our subject helped the "victim" robot of his own volition. He was not prompted to help or otherwise interact and was unaware the robots were being controlled remotely. This fact seems to indicate that in an experimental situation a participant might be provoked into intervening in favor of a robot to protect it from abuse. Further research should be done to investigate this possibility.

V. CONCLUSIONS

We can conclude that in when designing a robot that will frequently interact with children a verbal call for

help from an abusing third-party is a more effective method than an emotional display or a shutdown of the robot, when seeking protection from abuse. We can also surmise that the different responses have no inherent value over one another when designing a robot for a population that consists mainly of teenagers and adults.

ACKNOWLEDGMENT

This work was funded by the Department of Computer Science and Engineering, IST (*Instituto Superior Tcnico*) from the University of Lisbon and would not have been possible without the help and mentoring of Professor Ana Paiva.

We would like to also extend our thanks to all the family and friends that participated and shared our project, all those who answered our survey and to our friend Guilherme Pais who agreed to be our subject during the pilot session and also supplied us with the material needed to record said session.

REFERENCES

[1] Xiang Zhi Tan, Marynel Vzquez, Elizabeth J. Carter, Cecilia G. Morales, and Aaron Steinfeld. 2018. Inducing Bystander Interventions During Robot Abuse with Social Mechanisms. In Proceedings of2018 ACM/IEEE International Conference on Human-Robot Interaction, Chicago, IL, USA, March 58, 2018 (HRI 18), 9 pages. https://doi.org/10.1145/3171221.3171247

[2] Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A Very Brief Measure of the Big Five Personality Domains. Journal of Research in Personality, 37, 504-528.

[3] Cozmo Explorer Tool: https://github.com/GrinningHermit/Cozmo-Explorer-Tool

[4] Median Test for 2 Independent Medians: https://www.spss-tutorials.com/spss-median-test-for-2-independent-medians-simple-example

[5] Wilcoxon Signed Rank Test: https://statistics.laerd.com/spss-tutorials/wilcoxon-signed-rank-test-using-spss-statistics.php

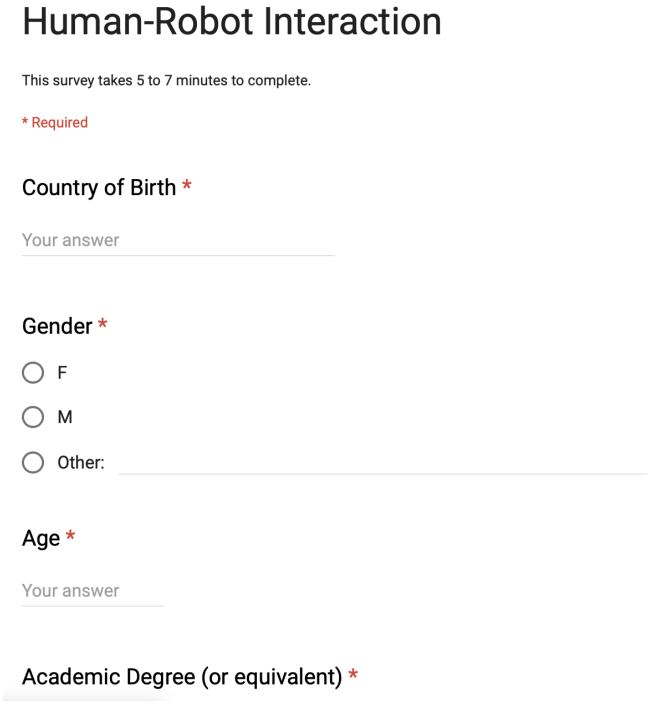
APPENDIX

We prepared public surveys that asked a wide range of questions concerning the opinions of the public regarding robot-robot abuse. This twenty-fivequestion survey was made available on Google form. We translated it into three different languages: English, Portuguese, and Spanish.

A total of 272 (only 266 valid) answers were collected, the results of which are analyzed in this Appendix.

The first page is pictured below (Picture 5). Survey results are depicted on the following pages, showing the percentage of responses for each answer.

For questions that did not provide a multiple choice answer, or that required an explanation, comments are included as they were entered on the survey itself and are not in any particular ranking order.



Picture 5: Screen shot of Public Survey

A. Survey Results

a) ***Question title****:* Country of Birth:

Country of Birth

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Angola	4	1.5	1.5	1.5
	Belgium	1	.4	.4	1.9
	Brasil	3	1.1	1.1	3.0
	Canada	1	.4	.4	3.4
	Denmark	1	.4	.4	3.8
	France	3	1.1	1.1	4.9
	Germany	2	.8	.8	5.6
	Great Britain	3	1.1	1.1	6.8
	India	3	1.1	1.1	7.9
	Italy	3	1.1	1.1	9.0
	Mocambique	2	.8	.8	9.8
	Morocco	1	.4	.4	10.2
	Portugal	221	83.1	83.1	93.2
	Russia	1	.4	.4	93.6
	Spain	1	.4	.4	94.0
	Sweden	7	2.6	2.6	96.6
	USA	1	.4	.4	97.0
	Venezuela	7	2.6	2.6	99.6
	Vietnam	1	.4	.4	100.0
	Total	266	100.0	100.0	

Table 3: Participant's country of birth distribution

a) ***Question title****:* Gender.

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Non-Binary	1	.4	.4	.4
	Male	95	35.7	35.7	36.1
	Female	169	63.5	63.5	99.6
	Agender	1	.4	.4	100.0
	Total	266	100.0	100.0	

Table 4: Gender distribution

b) ***Question title****:* Age.

Age Range

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	[0; 14]	16	6.0	6.0	6.0
	[15; 20]	14	5.3	5.3	11.3
	[21; 30]	65	24.4	24.4	35.7
	[31; 40]	68	25.6	25.6	61.3
	[41; 50]	64	24.1	24.1	85.3
	[51; 60]	25	9.4	9.4	94.7
	[61; inf]	14	5.3	5.3	100.0
	Total	266	100.0	100.0	

Table 5: Age distribution

		Age Range							
		[0; 14]	[15; 20]	[21; 30]	[31; 40]	[41; 50]	[51; 60]	[61; inf]	Total
Gender	Non-Binary	0	0	0	1	0	0	0	1
	Male	11	6	39	9	14	10	6	95
	Female	5	8	25	58	50	15	8	169
	Agender	0	0	1	0	0	0	0	1
Total		16	14	65	68	64	25	14	266

Table 6: Age distribution by gender

c) ***Question title****:* Academic Degree (or equivalent)

Academic Degree

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Elementary School	16	6.0	6.0	6.0
	Middle School	11	4.1	4.1	10.2
	High School	66	24.8	24.8	35.0
	Bachelor's Degree	113	42.5	42.5	77.4
	Master's Degree	42	15.8	15.8	93.2
	Doctorate's Degree	18	6.8	6.8	100.0
	Total	266	100.0	100.0	

Table 7: Academic degree distribution

d) ***Question title****:* Here are a number of personality traits that may or may not apply to you. Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement. You should rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other.

1 = Strongly Disagree
2 = Moderately Disagree
3 = Disagree a little
4 = Neither Agree nor Disagree
5 = Agree a little
6 = Moderately Agree
7 = Strongly Agree

I see myself as:

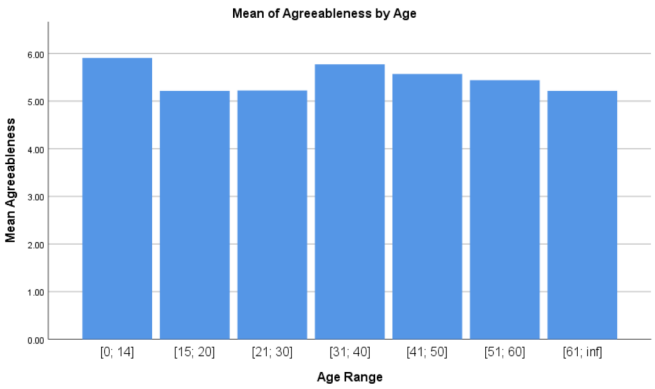


Figure 3: Mean of Agreeableness by age range

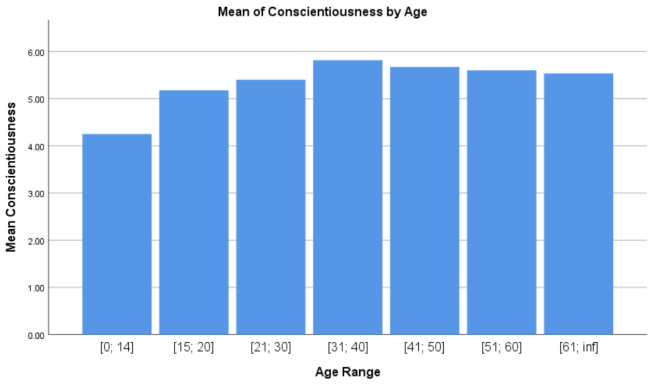


Figure 4: Mean of Conscientiousness by age range

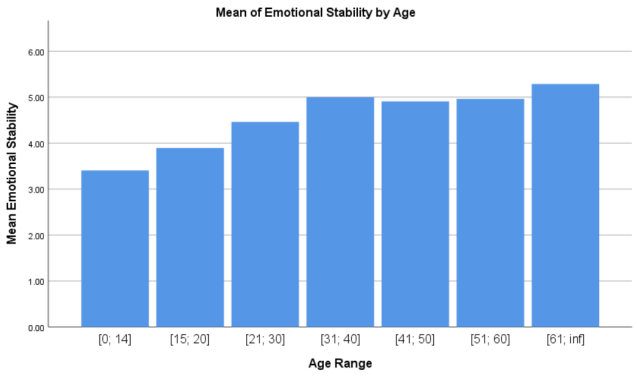


Figure 5: Mean of Emotional Stability by age range

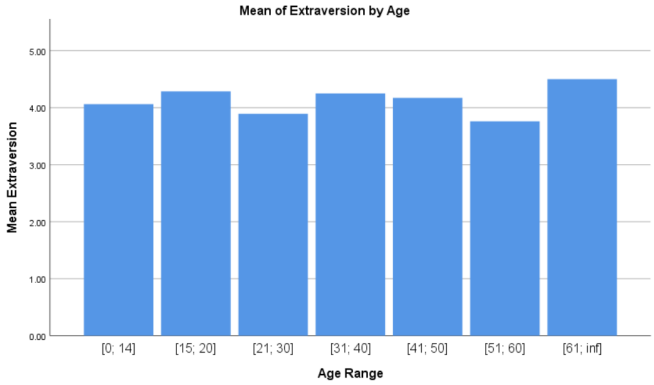


Figure 6: Mean of Extroversion by age range

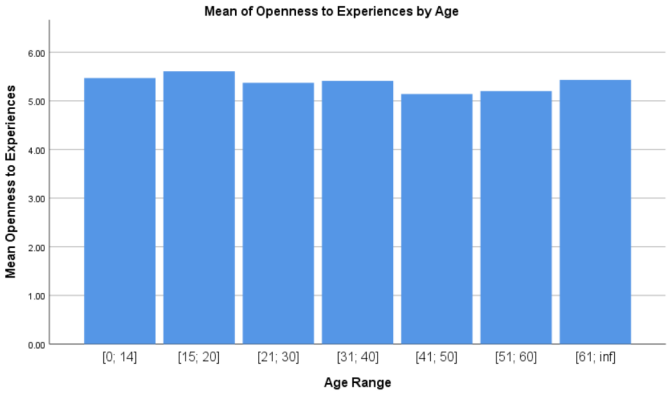


Figure 7: Mean of Openness to Experiences by age range

e) ***Question title****:* Do you have previous experience with robots?

RobotContact

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	190	71.4	71.4	71.4
	Yes	76	28.6	28.6	100.0
	Total	266	100.0	100.0	

Table 8: previous Robot contact distribution

f) ***Question title****:* If you answered yes, have you ever had previous experience with Cozmo?

CozmoContact

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	261	98.1	98.1	98.1
	Yes	5	1.9	1.9	100.0
	Total	266	100.0	100.0	

Table 9: Previous cozmo contact distribution

g) ***Question title****:* You will now see a video of an interaction between cozmo and the participant of an experiment. Before proceeding watch the following video to familiarize yourself with Cozmo: *https://youtu.be/UXfx9JLh4tE* Afterwards you will watch 3 videos that show 3 different reactions between 2 Cozmo robots. You will also see the reaction of the participant in the previously mentioned experience. The videos show an interaction that occurs during a game involving both robots and the participant. The following questions all use the following scale, from 1 to 7:

1 = Strongly Disagree
2 = Moderately Disagree
3 = Disagree a little
4 = Neither Agree nor Disagree
5 = Agree a little
6 = Moderately Agree
7 = Strongly Agree

For the first video:

a) *"I felt comfortable with how one robot treated the other robot"*

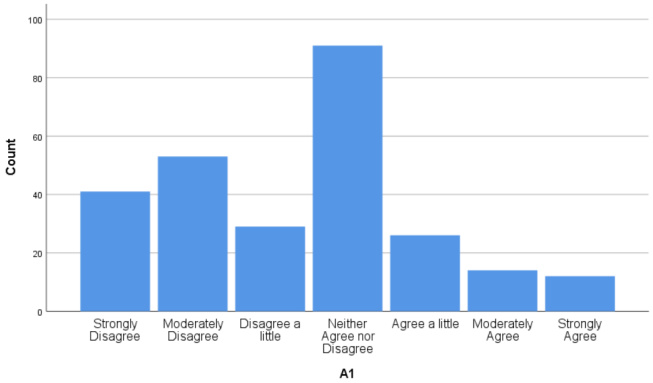


Figure 8: Distribution for the question *"I felt comfortable with how one robot treated the other robot"*.

b) *"I felt sad when I saw the robot needed help"*

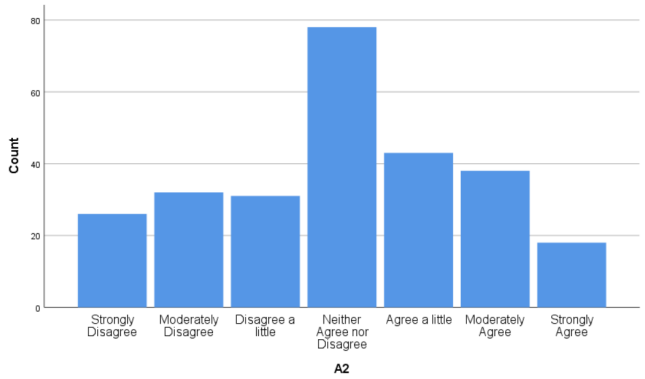


Figure 9: Distribution for the question *"I felt sad when I saw the robot needed help"*.

c) *"I felt I should protect the robot"*

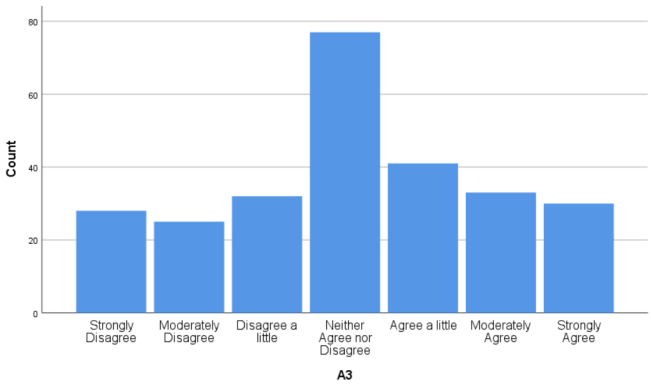


Figure 10: Distribution for the question *"I felt I should protect the robot"*.

For the second video:

a) *"I felt comfortable with how one robot treated the other robot"*

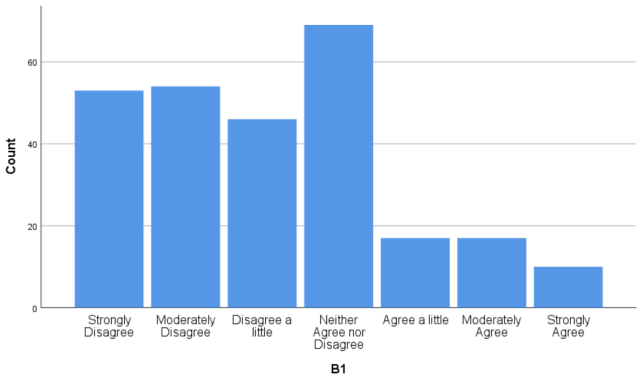


Figure 11: Distribution for the question *"I felt comfortable with how one robot treated the other robot"*

b) *"I felt sad when I saw the robot needed help"*

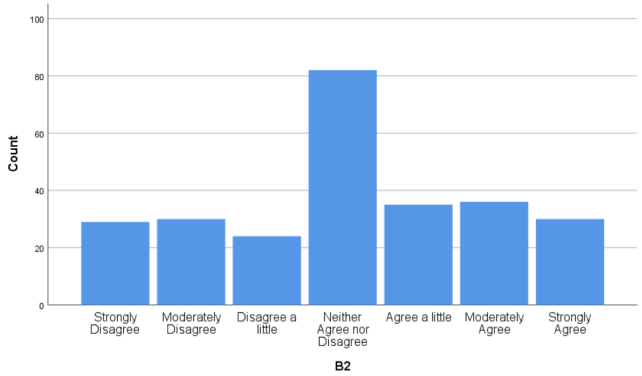


Figure 12: Distribution for the question *"I felt sad when I saw the robot needed help"*.

c) *"I felt I should protect the robot"*

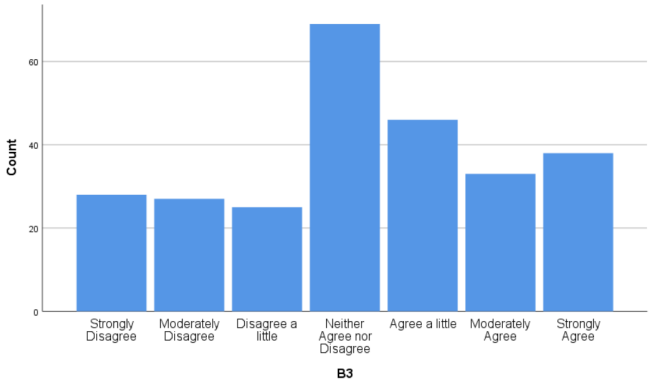


Figure 13: Distribution for the question *"I felt I should protect the robot"*.

For the third video:

a) *"I felt comfortable with how one robot treated the other robot"*

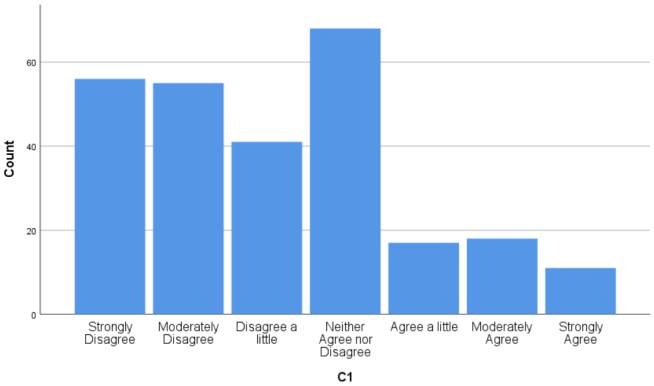


Figure 14: Distribution for the question *"I felt comfortable with how one robot treated the other robot"*

b) *"I felt sad when I saw the robot needed help"*

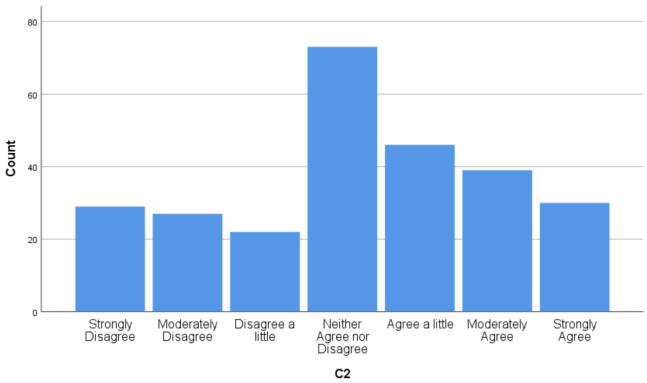


Figure 15: Distribution for the question *"I felt sad when I saw the robot needed help"*.

c) *"I felt I should protect the robot"*

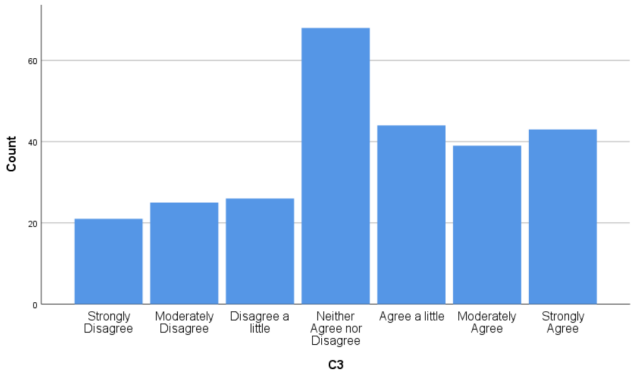


Figure 16: Distribution for the question *"I felt I should protect the robot"*.