Does the Safety Demand Characteristic Influence Human-Robot Interaction?

Jamie Poston(✉), Houston Lucas, Zachary Carlson, and David Feil-Seifer

Robotics Research Lab, University of Nevada, Reno 89557, USA
{jposton,houstonlucas,zack}@nevada.unr.edu, dave@cse.unr.edu

Abstract. While it is increasingly common to have robots in real-world environments, many Human-Robot Interaction studies are conducted in laboratory settings. Evidence shows that laboratory settings have the potential to skew participants' feelings of safety. This paper probes the consequences of this Safety Demand Characteristic and its impacts on the field of Human-Robot Interaction. We collected survey and video data from 19 participants who had varied consent forms describing different levels of risk for participating in the study. Participants were given a distractor task to prevent them from knowing the purpose of the study. We hypothesized that participants would feel less safe with the changed consent form and that participants' views of the robot would change depending on the version of consent. The results showed that features of the robot were viewed by participants differently depending on the perceived risks of participating in the study, warranting further inspection.

1 Introduction

The body of Human-Robot Interaction (HRI) knowledge is growing every day; however, the possibly confounding factor of the Safety Demand Characteristic (SDC) is not often recognized. The SDC is a demand characteristic that makes experiment participants in a laboratory setting feel more safe than in a real-world setting. Many studies in the field of HRI are often set in a laboratory setting because of ease and accessibility to researchers. This may affect the perceptions and actions of the participants. Due to the SDC, the experimental setting creates a sense of implied safety that may not exist in the real world. If the SDC alters the behavior of a person interacting with the robot, the results of many prior HRI studies may not translate well into the real world. For example, consider the effects the perceived safety of a robot might have on robot hand-offs [1,6]. A participant may be more willing and less hesitant to take an object from a robot in a laboratory setting. They may feel that the researchers will stop the study if something were to go wrong, or if the robot malfunctioned. In the real world, people do not have an inherent guarantee from researchers that they will not be harmed, and will be more likely to keep their safety in mind.

In this paper, we compare participants' perceptions of a robot after consenting to participate in a HRI study with either a standard consent form or an altered consent form that greatly exaggerated the amount of risk they assume

A. Agah et al. (Eds.): ICSR 2016, LNAI 9979, pp. 850–859, 2016.
DOI: 10.1007/978-3-319-47437-3_83

by participating in the study. First, we define the SDC and what it entails. Then, we present a controlled study in a laboratory setting that examines the affect of the SDC on participants' perceptions of the robot. We then present an analysis of these data.

2 Background

The "Safety Demand Characteristic" is a demand characteristic associated with participants in a controlled experiment environment [5]. Participants in a controlled experiment setting tend to feel inherently safe. They believe that the experimenter will not put them in harm's way, and the experimenter will not allow them to do any harm. This is a definite confounding variable, but the emergent field of Human-Robot Interaction does not explicitly account for this in the majority of associated research. If a human were to interact with a robot in the real world, the human may have a perception that is significantly influenced by caution. This may affect the applicability of HRI knowledge acquired in laboratory setting to the real world.

Studies of SDC found that when participants were asked if they would do a dangerous or harmful task, they would vehemently deny that they would do such a thing. However, when participants were asked by an experimenter to do those tasks, they comply because of the controlled experiment setting [7]. This shows how the laboratory setting affects how people act, even when they believe they would act a different way. In many current works, the field of Human-Robot Interaction generally does not take this factor into consideration, even though this could have an affect on participants actions and perceptions.

3 Methods

This section presents an experiment that observes how individuals interacting with robots feel when different levels of risk were communicated to them.

3.1 Robot Behavior

In this experiment, participants were asked to evaluate the robot's performance cleaning a table (see Fig. 1). This was a distractor task so that the participants would not know the true intent of the study. While the participants watched the robot, the robot would clean a section of the table and move closer and closer to the participant. Participants reactions to this were measured through surveys to determine if there was a difference between the two consent forms.

The robot began its motion when the experimenter left the room. It would look at the spot it was going to go, then clean that area using a back and forth arm motion. It did this four times, then looked sideways in the general direction of the participant and looked back at the table. It then translated sideways,

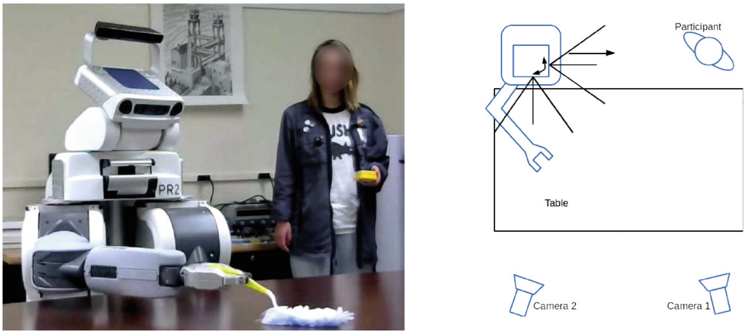

Fig. 1. The experiment set-up. Participants were asked to observe the robot as it cleaned the table. *Left:* The PR2 robot and participant (holding the E-STOP). *Right:* Top-down view of the experiment setting.

without changing its orientation to the table. It repeated cleaning three times, after which an operator in another room stopped the robot script.

To ensure experiment consistency, all of the robot actions were pre-scripted and autonomous. This removed the possibility of error from a human operator as well as the possibility of error or inconsistency from more complex autonomous behavior. The focus of this study was not to create a table-cleaning robot, but to investigate the SDC in regards to human-robot interaction. While the robot behavior could have been autonomous, we felt that a pre-scripted motion would ensure a consistent experience for all participants.

3.2 Experiment Manipulation

In order to change the level of risk perceived by the participant, we used two different consent forms. We used a between-participants design where participants individually participated in one of the two conditions. Participants would either be read a **standard** consent form that correctly enumerated the minimal risk of the study or an **altered** consent form that greatly exaggerated the risk of participating in the study. For the exact phrasing of the potential risk see Table 1. To make sure that the participants understood each aspect of the form and that they would not gloss over the changed variable, the consent form was read out loud to them before the study.

The independent variable was the assessment of risk in the consent form. The dependant variables included the participants reaction to the robot and the participants perceptions of the robot. We chose not to include a human control condition, where the robot is replaced by a human with the same actions, as there is no inherent danger with another person, but there could be with a robot.

Table 1. Phrasing for the standard and altered consent forms

Condition	Phrasing
Standard	If you agree to be in this study, you will be in a minimal risk setting
Altered	WARNING: If you agree to be in this study, you may be subject to physical harm or injury from the robot if proper caution is not used when interacting with the robot

Our hypotheses were:

H1: Participants will feel less safe with the changed consent form.

H2: Participants' views of the robot will change depending on the version of consent.

The first hypothesis addresses the core aspect of the study, that participants can feel unsafe in a laboratory setting. The second hypothesis deals with the correlation between safety and positive perceptions of the robot.

3.3 Experiment Protocol

The consent process took place in a separate room from the robot. During the consent process, the consent form was read out loud to the participant to guarantee that they understood each part. After the consent forms were read and signed, the participants were led into the room with the robot. They were told that they were to be evaluating the PR2's performance when cleaning a table. The participants were given an e-stop and instructed on its use, including that they should press it if they felt unsafe and the robot would stop. The experimenter then left the room, and the robot cleaned the table for 2 min. The robot moved towards the participant in this time so as to increase the perceived risk of the whole scenario. The participants were not told that the experimenters were still able to see in the room through a web-camera, so as not to alter the affect of being alone with the robot. After the cleaning task was done, the participants were led into a different room to take the survey.

The PR2 robot was chosen for its mobility and size. The PR2 is able to translate sideways along the table, as opposed to turning and moving like other robots. The relative size of the PR2 as compared to a human is significant as it creates a definite negative impact if the robot were to bump a human. If a robot were too small, it may not be perceived as able to do any harm to a human.

The robot's motions were pre-scripted so as to be the same for all participants. To clean the table, the robot moved its gripper holding a duster in a back and forth motion in four different areas, looking at each of the areas before it cleaned them. Before the robot moved towards the participant, it looked at them and looked back at the table. The room was laid out in a way that the participants could not move around the table or out of the way of the robot should it continue to move towards them (see Fig. 1).

After the robot was finished attempting to clean the table, we asked participants to complete a survey of their perceptions of the agent during these activities. The participants were led to a different room to complete the online survey on a computer. After they finished the survey, each participant was then debriefed on the deceptions of the study, including the true title and nature of the study and the altered consent form. They were then given a copy of the standard consent form to keep for future reference.

3.4 Participant Recruitment

Participants were recruited by word of mouth at University classes and clubs individually. They were then scheduled to meet at an appointed time to participate in the study. We collected data from a total of 19 participants, 9 male, 8 female, 2 reported nonbinary gender. The majority of the participants were between the ages of 18 and 28 years old, with an average age of 21 years old with a standard deviation of 3 years.

Not included in the total count of the participants were 4 participants who had different experiences than the rest. Two of these participants had the robot fail to complete its cleaning task during the study. One of the participants discovered the purpose of the study, and thus may have been biased when filling out the survey. And the last participant was the only participant to hit the e-stop button. While hitting the e-stop was something that we had expected, since no other participants did the same we elected that their experience was significantly different than the other participants, since the robot was less than halfway through with the cleaning cycle.

3.5 Data Collection

We used an online survey using Google Forms to record quantitative and qualitative responses, as well as demographic information. We asked a total of 41 questions in 8 sub-scales. The sub-scales that covered participants' perception of the robot were from the Godspeed Questionnaire [2]. The other sub-scales that covered comfort and trust in the robot were taken from the Negative Attitudes Towards Robots (NARS) survey [3,4,9]. These sub-scales did not include a part of the survey that asked the participants to evaluate how well the robot cleaned the table. Between the altered and standard consent form conditions, the questions were kept identical. Out of the 41 questions, 39 questions were on a scale of 1 to 5 and 2 questions were free-response questions that allowed the participant to state the positive and negative aspects of the robot's behavior. Participants filled out the survey on a computer in a room separate from the robot.

In an effort to better understand how the inherent feeling of safety in a controlled setting affected participants, we measured participants responses in 8 different categories: Anthropomorphism, Animacy, Perceived Intelligence, Perceived Safety, Robot Trust, Comfort in Setting, Predictability, and Dependability. For more detail about these measures, see Table 2.

Table 2. Categories and number of questions from online survey provided to participants. All questions were on a 1–5 Likert scale

Category	# Questions
Anthropomorphism (ANT)	5
Animacy (ANM)	5
Perceived Intelligence (PI)	5
Perceived Safety (PS)	3
Robot Trust (RT)	5
Comfort in Setting (COM)	4
Predictability (PRD)	6
Dependability (DEP)	2

We also recorded the behavior of the participants using a video-camera that we used to see if the participants behaved differently between conditions.

4 Results

To analyze the data we ran unpaired Student's t-tests between conditions in each category. No conditions showed significance at a ($p < 0.05$) level. There was a non-significant drop in the values of Intelligence, Comfort, and Dependability (see Fig. 3).

Table 3. Pearson's r values for every category.

	ANT	ANM	PI	PS	RT	COM	PRD	DEP
Anthropomorphism	1.000	0.370	0.376	0.159	0.613	0.520	0.653	0.422
Animacy	0.370	1.000	0.242	0.152	0.342	0.077	0.443	−0.071
Perceived Intelligence	0.376	0.242	1.000	−0.311	0.583	0.303	0.511	0.520
Perceived Safety	0.159	0.152	−0.311	1.000	−0.255	−0.221	−0.307	−0.362
Robot Trust	0.613	0.342	0.583	−0.255	1.000	0.570	0.730	0.761
Comfort	0.520	0.077	0.303	−0.221	0.570	1.000	0.274	0.782
Predictability	0.653	0.443	0.511	−0.307	0.730	0.274	1.000	0.470
Dependability	0.422	−0.071	0.520	−0.362	0.761	0.782	0.470	1.000

To check for correlation between safety and positive ratings of the robot we used Pearson's r. Safety showed no strong correlation with any of the robot ratings. The specific Pearson's r values can bee seen in Table 3. No significant correlation was found between self reported safety and any other conditions.

Post-hoc analysis did show strong positive correlation of Dependability with both Comfort ($r = 0.782$) and Robot Trust ($r = 0.761$) (see Fig. 2).

To analyze behavioral data between groups, two independent raters coded mutually exclusive behaviors with the recorded video. We report the behaviors

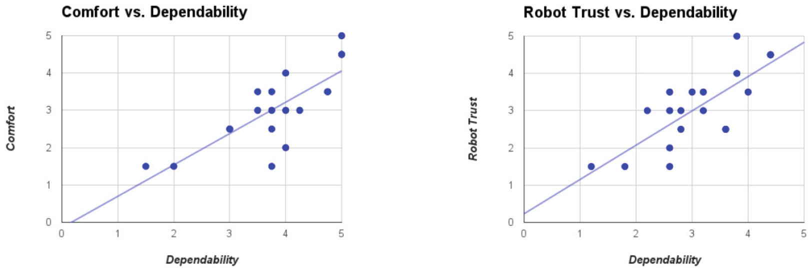

Fig. 2. *Left:* Comfort vs Dependability Correlation: Pearson's Product = 0.782; *Right:* Robot Trust vs Dependability Correlation: Pearson's Product = 0.761

(Leaned Closer to Robot and Looked at Video Camera) where the raters had high agreement (Cohen's-$\kappa > 0.60$). The behavior 'Looked at the Video Camera' showed weak significance ($\chi^2(1, N = 19), p < 0.1$). The behavior 'Leaned Closer to Robot' strong significance ($\chi^2(1, N = 19), p < 0.05$). In both of these categories the participants in the Altered condition were more likely to exhibit these behaviors.

The participants were provided a free-response section where they were asked what they disliked about the robot as well as what they liked. Here are some typical responses:

Liked:

- "I liked how it checked where it was going before it moved, and how it checked the table for thoroughness." -A4
- "The way it looked up and around before moving." -B2

Disliked:

- "I have to admit that I was slightly uncomfortable with how close it was getting but that is the same with humans that are doing something and I can't get out of the way."[sic] -A5
- "It also would look at me before it moved towards me in order to clean more of the table. This was rather unsettling because it made me feel as though it was making the decision to move closer to me."[sic] -B3

5 Discussion

The goal of this paper is to explore the impact of the implied safety of a laboratory setting on perceptions of Human-Robot Interaction. The study results do not support **H1**. Participants did not report feeling less safe with the altered consent form. However, this seems to be due to a uniformly low rating of safety, regardless of experiment condition. This suggests that the robot behavior of gradually moving closer to the participant was not comfortable to the participant

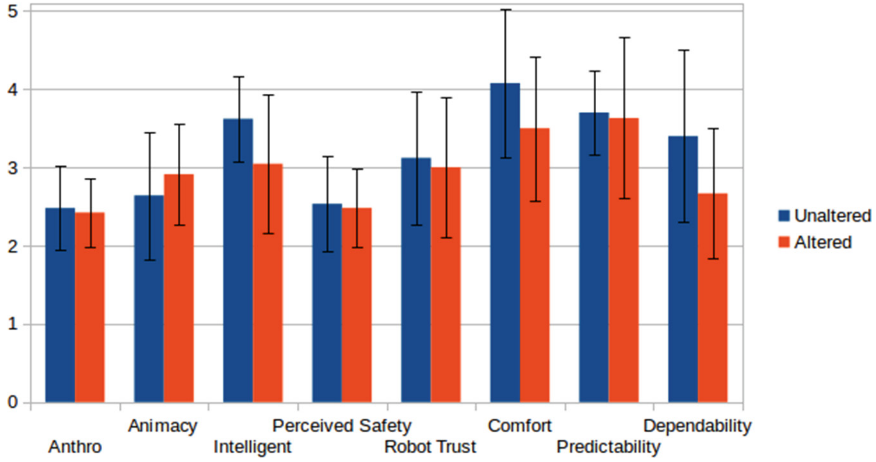

Fig. 3. Category Means: No significance at the $p < 0.1$ level found.

and the data may be showing a floor effect. There was no significant correlation between the self-reported feeling of safety and the rating of the robot.

The results presented in the previous section partially support **H2**. Looking at the free-response questions, the differences in how users perceived the robot's behavior from the standard to the altered consent conditions suggest that communicated risk might affect how a participant perceives the robot. What is interesting with the free-response questions is that the same robot actions were perceived very differently when the communicated risk was greater. The changes in ratings of the robot's intelligence, comfort, and dependability (though not significant) also indicate that the perceptions of the robot's behavior changed when increased risk was communicated.

Furthermore, the behavioral analysis suggested a change in perception of the robot. Participants in the altered consent condition frequently looked at the camera or leaned in to look at the robot. This may suggest an expectation of the participants that the experimenter would step in and stop the experiment if necessary. It may also be that leaning allowed the participants a better view of the robot while also affording the chance to stay further away from the robot. As these behaviors indicate a greater feeling of discomfort, these behaviors partially support **H1**.

A possible explanation for the lack of difference in perceived safety is that there were other effects that are skewing the data. One such effect could be that the participants' self reported safety was measured by a survey that took place after their interaction with the robot. Since the participants filled out the survey after the perceived danger had already passed, their perception of danger was in hindsight when they know that no harm had come to them. This hindsight could be biasing their responses.

A limiting factor of our study was that it had only 19 participants. This may not have been a large enough participant pool to get viable data. Also, the setting of the study may have biased the participants. While the study was held in a laboratory, there were other robots in the same room as the participants when they completed the survey. The main robot was not there, but this may have prompted participants to compare the other robots in the room they took the end survey with the robot that they had interacted with.

Some of the participants knew the proctor or someone else in the room where the surveys were administered. This may have affected their feeling of safety and comfort, as they may have assumed that their prior relationship would gain them special considerations and protection.

6 Conclusion and Future Work

In this study, the researchers postulated that participants would feel less safe with the changed consent form and that participants who felt safe would also feel more positively about the robot. Data from 19 participants were collected via an online survey, video analysis, and free responses. Statistical analysis on these data showed some significant differences between the altered and standard consent form conditions. A post-hoc analysis showed positive correlation between comfort and trust in the robot that varied with the perceived dependability of the robot. The changes in participant behavior toward the robot, and the perception of the robot's actions indicate the potential for follow-up on **H1** and **H2**. Follow-up research and reconfiguration of the study may further support the experimental hypotheses provided.

While this exploration was not able to conclusively give evidence for the SDC, it does suggest an effect due to communicated risk of a laboratory study. This opens up a wide area of future work considering how perceived risk might affect HRI studies conducted in laboratory settings. In this study, the participants had the e-stop the entire time, the robot looked at them before moving every time, and the robot was a consistent, moderate speed in both conditions. Any one of these variables could have had a more significant effect on the participants' perceived safety and perceptions of the robot.

Gaze has been shown in the past to be able to influence participants' perceptions of the robot [8]. Not only that, but gaze can directly affect participants' actions and strategy in studies examining object hand-offs [1,6]. The e-stop was explained to the participants as being able to stop the robot if they felt unsafe. This may have contributed an inherent feeling of safety for the participants, and the researchers will examine how the participants would react without the e-stop. The robot also moved towards the participants at a steady, moderate pace. If the robot was more abrupt and quick about moving, this might provoke a greater response in the participants.

The researchers plan on continuing this work with what they learned from this study in order to further examine the effect of perceived safety on participants' actions and perceptions of the robot. Specifically looking into the effects that the gaze of the robot, possession of the e-stop, and speed of the robot has on the safety perceived by the participants. Further work is needed to fully examine the causes and effects of the SDC.

Acknowledgments. This material is based upon work supported by the National Aeronautics and Space Administration under Grant #NNX10AN23H issued through the Nevada Space Grant, the Office of Naval Research DURIP award #N00014-14-1-0776, the National Science Foundation #IIS-1528137, and the UNR NSF EPSCoR UROP Program #IIA-1301726. We appreciate the help from Mercedes Anderson, Gaetano Evangelista, and Nathan Yocum who helped administer the study.

References

1. Admoni, H., Dragan, A., Srinivasa, S.S., Scassellati, B.: Deliberate delays during robot-to-human handovers improve compliance with gaze communication. In: Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction, pp. 49–56. ACM (2014)
2. Bartneck, C., Croft, E., Kulic, D.: Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. Int. J. Soc. Robot. **1**(1), 71–81 (2009)
3. Kraft, K., Smart, W.D.: Seeing is comforting: effects of teleoperator visibility in robot-mediated health care. In: The Eleventh ACM/IEEE International Conference on Human Robot Interaction, HRI 2016, pp. 11–18. IEEE Press, Piscataway (2016). http://dl.acm.org/citation.cfm?id=2906831.2906836
4. Lee, J., Moray, N.: Trust, control strategies and allocation of function in human-machine systems. Ergonomics **35**(10), 1243–1270 (1992). doi:10.1080/00140139208967392
5. Martin, D.: Doing psychology experiments. Cengage Learning (2007)
6. Moon, A., Troniak, D.M., Gleeson, B., Pan, M.K., Zheng, M., Blumer, B.A., MacLean, K., Croft, E.A.: Meet me where i'm gazing: how shared attention gaze affects human-robot handover timing. In: Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction, pp. 334–341. ACM (2014)
7. Orne, M.T., Holland, C.H.: On the ecological validity of laboratory deceptions. Int. J. Psychiatry **6**(4), 282–293 (1968)
8. Plaisant, C., Druin, A., Lathan, C., Dakhane, K., Edwards, K., Vice, J., Montemayor, J.: A storytelling robot for pediatric rehabilitation. In: Proceedings of the Fourth International ACM Conference on Assistive Technologies, pp. 50–55. Arlington (2000)
9. Yagoda, R.E., Gillan, D.J.: You want me to trust a robot? The development of a human-robot interaction trust scale. I. J. Soc. Robot. **4**(3), 235–248 (2012). http://dblp.uni-trier.de/db/journals/ijsr/ijsr4.html#YagodaG12