Good Robot Design or Machiavellian? An in-the-wild robot leveraging minimal knowledge of passersby's culture

Elaheh Sanoubari[1], Stela H. Seo[1], Diljot Garcha[1], James E. Young[1], Verónica Loureiro-Rodríguez[2]
[1]Department of Computer Science, [2]Department of Linguistics, University of Manitoba, Canada
{e.sanoubari, stela.seo, garchads, young}@cs.umanitoba.ca, v_loureiro-rodriguez@umanitoba.ca

***Abstract*—Social robots are being designed to use human-like communication techniques, including body language, social signals, and empathy, to work effectively with people. Just as between people, some robots learn about people and adapt to them. In this paper we present one such robot design: we developed Sam, a robot that learns minimal information about a person's background, and adapts to this background. Our in-the-wild study found that people helped Sam for significantly longer when it adapted to match their background. While initially we saw this as a success, in re-considering our study we started seeing a different angle. Our robot effectively deceived people (changed its story and text), based on some knowledge of their background, to get more work from them. There was little direct benefit to the person from this adaptation, yet the robot stood to gain free labor. We would like to pose the question to the community: is this simply good robot design, or, is our robot being manipulative? Where does the ethical line lay between a robot leveraging social techniques to improve interaction, and the more negative framing of a robot or algorithm taking advantage of people? How can we decide what is good here, and what is less desirable?**

Keywords—social robots, persuasive robots, culture, in the wild

I. Introduction

Robots can nowadays be found in public spaces such as airports, shopping malls, museums, or hospitals, where they interact with the general public. In these contexts, robots are commonly designed to leverage people's existing social interaction skills. These social robots often have anthropomorphic or zoomorphic designs and use human-like gestures, speech, and behaviors to interact with people [70]. This can be quite positive, as social interaction is fundamental for user comfort [19] and, if done well, can help a robot be easy to understand [28]. There are also potentially negative cases, such as social robots designed to make a person feel bad [55], or persuasive robots that leverage social interaction to pressure people to do things they may rather not do or may not be in their best interest [5,12,21,47]. In many cases, however, it is not clear where a social robot interaction design falls on this spectrum.

In this paper, we present a case study from our own work which we believe brings up a discussion on some of the ethical implications of social robot interaction design. We designed, implemented, and deployed a robot (named Sam) in-the-wild. Sam uses a simplistic cultural model to adapt to users, where it asks a simple question and then uses a targeted script (back-story, motivations, word choices, etc.) based on where the participant is originally from. Sam asks people to help it with a task, which continues indefinitely until the person elects to quit. We note that the task is not enjoyable and there is no direct benefit to the person. Our study results found that participants helped Sam for ~30% longer when Sam matched participants using its culture-based model (average 5min 34sec), in comparison to when Sam mismatched (average 4min 22sec). Thus, from asking only a single question of participants and simply changing a script, our robot was able to get people to help it for significantly longer. Our express design goal with Sam was to leverage minimal knowledge of user background to improve interaction with them.

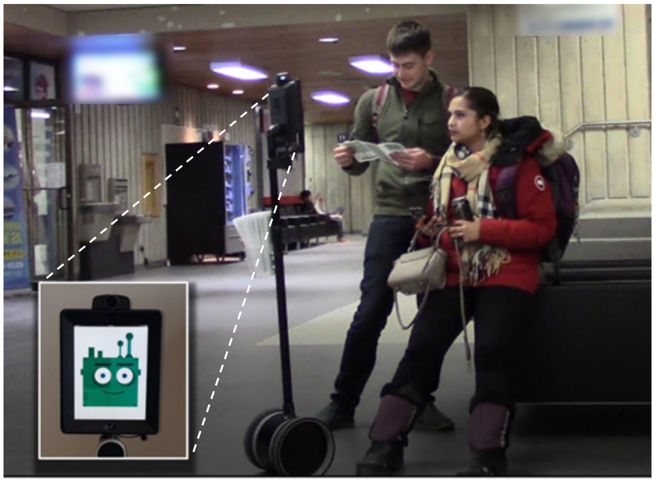

Fig. 1. Participants interacting with Sam in a public space at our university. Sam adapts to the user's background to get people to help it for longer.

On the one hand, our work provides a novel, simple-to-implement social interaction technique that can help robots in the wild be more effective at getting their task done. On the other hand, in retrospect we considered that perhaps this is an illustrative example of just how easy it is for a robot to manipulate people: simply by changing a script to match people (effectively, lying about its motivation and group relations), our robot gained more free labor from unsuspecting people. In looking back at our study, we ourselves have struggled with this question, of whether Sam is simply good effective social robot design, or whether it is being devious.

Stepping back, this work has caused us to more broadly consider the question of social robotics in general. Where is the line between a robot leveraging, e.g., a smile to be helpful and create comfort, or a carefully-timed calculating smile to manipulate and take advantage of people? Or are both okay?

In our exploration we aim for practical considerations and questions facing roboticists in their daily work designing and deploying social robots. We present our novel adaptive robot design as a case study that brings up a larger problem in HRI, which is the ethics of using social interaction techniques to change human mood and behavior. We invite the reader to consider whether they would feel comfortable with our design (and more advanced versions) in public spaces. We hope that this process will generate discussion and debates within the community, and finish the paper with a series of pragmatic discussion points emerging from our work, to serve as a springboard for this conversation.

II. Related Work

It has been well established in the literature that social robots can be potentially used for the detriment of people. Some robots are purposefully designed in both appearance and behavior to exert power over people, for example being placed in a position of authority [5], or by giving commands [21]. Such robots can employ manipulative or coercive social behavior to pressure people to comply to requests that may be uncomfortable or not in their best interests [5,21]. This behavior can be fairly subtle, such as robots designed to build emotional bonds, only to exploit them for security gains [15]. A related recent study illustrated how a robot, disguised as a food delivery bot, persuaded people to let it gain access to a secure facility [12]. Here the undesirable, perhaps nefarious application of social robots is quite evident in the examples: robots can use social interaction techniques to manipulate, pressure, and even trick people into doing things which are not in their best interests. Our robot Sam, is much more nuanced, and serves as an exemplar of robot techniques which are not so easily categorized in a negative fashion.

There are many such robots and techniques. For example, some designs aim to use social interaction techniques to modify human behavior for an ostensibly desirable goal, such as to encourage energy savings [41] or to support health management [37,38]. Robots as team members can purposefully shape and influence social dynamics in a group, and between human team members, to improve team effectiveness [33,53]. In these cases, it is easy to position the techniques in a positive light as good interaction design. However, we raise the question of whether it is indeed okay for a robot to be programmed to algorithmically manipulate and modify a person's behavior and their interactions with others, especially if the person involved is not aware of this robot goal. As demonstrated by our own project, such techniques can easily be used to the benefit of the robot, only, and not the user.

A key area of social robotics in this regard is their ability to induce emotional reactions in people, often through mechanisms relating to empathy. For example, a robot can concoct a scenario to make people feel bad on purpose [55], and can leverage this for its own goals, such as to discourage people from turning it off by begging [6], or even inducing empathy to appear more useful or increase peoples' willingness to use them [62]. Research has shown how people experience strong negative physiological reactions when seeing a robot hurt or abused [48,61]. People also experience positive reactions, for example, the Paro baby seal robot can help calm people and create a positive experience. [58]

This reaction, to respond and relate to apparent emotion in a robot, appears to be quite natural and even occurs in children. For example, children may adjust how they talk to a machine to avoid hurting its feelings, and their relationship with a robot can impact their self-confidence [63,64]. Emotional reactions may also be difficult to avoid, as even in professional contexts where people are well-trained, such as in the military, people experience empathy toward robots that shapes their relationships with them and impacts how they are used [20]. Empathy is commonly used in social robot design due to its effectiveness. In our work, we aim to leverage culturally-matched speech to increase empathy for the robot. However, in reflection we pose the question of whether it is acceptable for a robot to purposefully shape human mood and behavior.

Many have considered this broader question of the ethics surrounding social robotics techniques. This follows an established inquiry of "dark" interactions [22], where usability principles can be leveraged for the detriment of a user (and perhaps the benefit of a company or stakeholder). In robotics, negative uses of social robots has been coined "psychological attacks" [15,47], relating socially manipulative behavior to literature in security. We feel that our robot Sam is an important contribution to this space as an arguable case (as "dark", or not) that serves as an anchor point emphasizing just how nuanced and complex the issue of ethics surrounding social robots is.

There has also been reflection from the higher level, such as regarding the ethics surrounding having "inauthentic" social robots care for our children [63]; for example, children may be cruel to social robots without natural consequences typical when they are cruel to other people or animals. Similarly, people may develop inauthentic relationships with social care robots, which do not actually reciprocate caring and are only programmed to do so [56,65]. Sam falls clearly within this space, as it uses fake stories to build culture-targeted comradery, resulting in people helping it for a longer duration of time.

In the next component of this paper, we present our novel HRI robot design and experiment. We invite the reader to consider whether this is just effective interaction design or if our robot is manipulative, and what we as a community want to say about such robots. We then present a series of discussion points that we pose to the community surrounding this issue.

III. Sam: Adapting to Users to Get More Help From Them

We designed and implemented Sam, an in-the-wild robot that gleans knowledge of a person's cultural background, and uses this to decide how to interact with the person. Sam is an inverted pendulum robot with a computer-animated robotic face (Fig. 1), a unisex (mid-register) synthesized voice, and a unisex name. Sam's purpose is to enter public areas, engage in conversations with passersby and ask them to help it with a task. If someone agrees to help, Sam asks the person some basic questions, and then uses their answers to change its script to match their cultural background. The goal of this design is to be more relatable to people, to increase their empathy, and to encourage them to help the robot for a longer time.

We first present Sam, and the results from our study, and follow with our analysis on whether Sam's actions are acceptable. We do this on purpose, to present the work in a neutral fashion typical to human-robot interaction studies papers. However, we ask the reader to consider as they read the implications of Sam's interaction design and behavior.

A. Motivation and Approach

In human-human interaction, it is common for individuals to change how they communicate with others according to age, social rank, mood or the context of the interaction [4]. For example, when interacting with children, people tend to use simpler and friendly words, and to stoop down while talking. People also accommodate social norms of the country they are visiting; for instance, an American might bow instead of giving a handshake while in Japan. In a highly multi-cultural environment, such as a university campus, we may find ourselves adapting to language ability, cultural expectations, and religion, on a regular basis. People typically adapt to improve the quality of the interaction, for example, to be more liked, to avoid offending people, or to communicate more clearly. Further, even without adaption people tend to feel more positive ties and connections to people they view as similar to themselves (called homophily [40]). We note that robots likewise can change their communication style while interacting with people for similar goals.

We hypothesize that, in a multicultural environment (such as a university campus, or an airport), a robot that adapts to the cultural background of users may be able to be more effective in working with them. Specifically, we investigate the effects of a robot adapting to people's background on how willing passersby are to help the robot with a menial task.

B. Simple cultural Model

Culture can be defined as the attitudes and social norms common to a group of people [72]. It underlies various aspects of our social behaviors, affects our reasoning style [32], and shapes what we deem appropriate in our interactions with others. Culture varies wildly across the world and between social groups, and is recognized as a key variable in understanding social behavior [10,39].

The idea of culture is broadly encompassing and thus it can be difficult to succinctly define. Researchers have proposed a range of models and theories that break down various components of culture (e.g., see [14,25,30,60]). One prominent model is Hofstede's 6-D model of national cultures that focuses on social values, with six dimensions for description and analysis [30]: power distance, uncertainty avoidance, individualism versus collectivism, masculinity versus femininity, long versus short term orientation, and indulgence versus restraint. This model established baselines for describing cultures around the world using these dimensions. We base our work on this model to coarsely differentiate national cultures.

We narrow our focus to the individualism dimension for our work, in part as it has some of the strongest data support [31], but also as it is relatively easy to employ: the individualism dimension ranks national cultures on a continuum from individualist to collectivist, with countries being chiefly individualist or collectivist [30] (with few in between). This dimension draws geographical distinctions, with North American and European countries having higher individualism scores, and Asian, African and South American countries having lower scores (thus, being more collectivist). For example, the individualism score for United States is 91, while it is only 13 for Columbia and 6 for Guatemala [29]. While we accept that this is only a coarse-grained classification, a robot could categorize a person's back-ground as individualist or collectivist by finding out which country's culture they identify with.

In individualistic cultures people generally prefer to act on their own rather than as a group, directly protecting their own interests [24]. In contrast, people in collectivist cultures prefer to integrate into groups and mutually supportive relationships, where they support the group which in turn group protects them. Individualist societies tend to have looser in-group ties and form smaller groups, typically closely associating with immediate family members and few others, while individuals in collectivist society are expected to offer support to their extended family members when they face difficulties [24]. Independence is typically valued more in individualist societies while collectivists value interdependence more. [24]

Cultural background tends to influence communication style preferences [23], which correlate well with individualism: collectivist cultures tend to prefer implicit and indirect communication, and to avoid conflict, whereas individualist cultures prefer explicit and direct communication. [24]

Thus, a broad-brush culturally adaptive robot could first learn about which country a person is from, and then roughly categorize them as individualist or collectivist. Following, the robot could use language, back story, and communication styles more common to those cultural backgrounds, in an attempt to generate empathy and improve sense of relating to the robot.

Culture, in general, has been broadly studied in HRI. Work includes investigating attitudes toward robots across cultures (e.g., [8,27,42–44,46]), and uncovering culture-specific expectations and design preferences [35,36,45]. A person's culture impacts how their interactions with a robot unfold (e.g., [11,50,67,68]), for example, a person's culture can determine how a robot's communication style impacts credibility [49], or how they rate a robot's social acceptance, likeability and trustworthiness [3,42]. Furthermore, similar to people, the robot itself can act differently in different cultures, to impact how it is perceived [27,35] and how much people trust its opinion [3,67]. We build on this body of work by creating a robot that leverages knowledge of culture in the wild in an attempt to impact how much people help it.

C. Interaction Design: an In-the-wild Robot

While the majority of social HRI research has been carried out in a traditional lab setting [9], the merits of conducting experiments in ecologically valid, in-the-wild contexts are being increasingly recognized [34,52]. We can expect people to act differently in a natural setting compared to in-the-lab setting, and robot behavior to be more realistic in noisy real-world spaces. Recent studies have investigated interaction in pubic spaces including rural areas [16], shopping malls [1], cafés [2], and classrooms [66,69]. As such, we aim for our robot to interact with people naturally in a public setting.

Sam interacts verbally: it speaks to people and listens to their answers. It uses on-screen text to give instructions, but people answer verbally and never touch the robot or screen. Sam's only physical interaction is to approach people and maintain an interaction distance of ~1m. Sam has a narrow interaction tree: if participants change the topic, for example, by asking questions about the robot or making jokes, Sam simply states that it does not understand and re-iterates its last statement or question. As such, Sam has only minimal intelligence and follows a simple interactive behavior.

Sam initiates interaction by approaching a person passing by and verbally asking for help. If they agree, Sam starts by asking a few questions that it uses to culturally profile the person, and then proceeds to tell a quick backstory that highlights the robot's purposes and goals. Sam's stated purpose is to get help from people to improve its capabilities. We use a backstory to help encourage engagement and believability [59]. Sam then asks the person to help it and continues asking for help indefinitely. If at any point the person indicates that they want to leave (e.g., verbally, or by just leaving), Sam pleas for them to continue.

1) *Task*

Sam's self-stated goal is to get help from passersby to improve its capabilities. We selected image labelling, where the robot shows an image with a caption to the person and asks a question about its content, instructing them to answer verbally ("yes", "no", or "skip", Fig. 2).

We selected this as an easy task that people may understand to be difficult for a robot, particularly as many cases could be easily seen as a tricky (but still reasonable for a person). We compiled a database of about 2000 images to be labeled (from crowdsource.google.com), filtered for appropriateness and shuffled to prevent duplicates and for improved believability (i.e., to avoid seeing repeated cases between people).

2) *Identifying Participant Culture*

Sam attempts to identify a person's culture by asking them. However, to mask the question's purpose, the inquiry is situated within other questions (see the *Questions* panel of Table 1), framed simply as Sam wanting to learn about the person. We carefully considered the phrasing of this question. Given the multi-cultural nature of the university environment in Canada we could not ask where people were born, as they may have immigrated as a child, or where they have lived longest or in the recent years, as they may have attended international schools. Automated visual processing is challenging given the diverse phenotypes and clothing common in Canada. Asking people's own opinions on their cultural identity was a compromise that would help us generally classify people according to their cultural background.

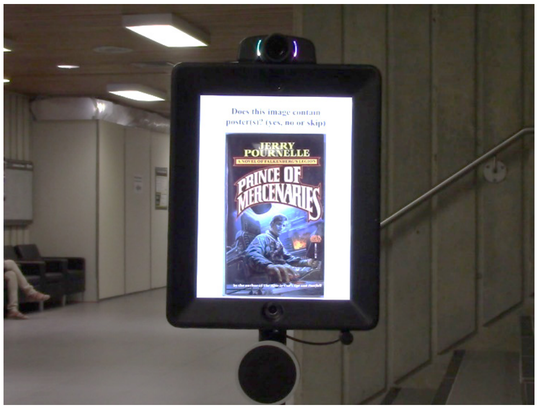

Fig. 2. Example of an image labeling question – Sam's face disappears during the task. Note the question at the top: "does this image contain poster(s)? (Yes, no or skip)".

Sam references a published corpus of how countries rank on the individualism dimension [29] and classifies a participant as being individualist if their cultural score is higher than 50 (on a 100 scale), and collectivist otherwise. As explained earlier, most cultures are predominantly collectivist or individualist.

3) *Manipulation: Interaction scripts*

Sam changes its script to align with people's culture to encourage them to relate to it and increase empathy, to ultimately get more help from the person. Once Sam infers background it selects either a collectivist or individualist script for the rest of the interaction to match the person. We designed these scripts in collaboration with an academic linguist, and balanced script timing, length, and variety. The full script is in Table 1, based on prior work. [54]

The individualist script emphasizes independence. For example, in the backstory Sam claims it needs help to become "more independent from others". During interaction, Sam says things like "your answers help me stand on my own feet!" When the person tries to end interaction, Sam's pleas include things such as "but how do I become independent without your help?".

In the collectivist script Sam emphasizes interdependence. For example, in the backstory Sam states that it needs help to be a more useful member of its team. During interaction, Sam says things like "Thank you! Your answers help me be more useful in my team." Sam pleas with people trying to end interaction by saying things such as "But how do I make my team proud without your help?"

4) *Implementation*

Sam is a Double telepresence robot with an Apple iPad Air 2 for a face. Sam is remotely operated using the Wizard of Oz method [51], although people are led to believe it is autonomous.

IV. In-the-wild Experiment

We conducted a between-subjects, in-the-wild experiment to explore the effects of a robot changing its script based on a passerby's culture on how long a person is willing to help the robot. As a base case we contrasted matching participant culture against intentionally mismatching it. We decided against attempting a *culturally neutral* condition given the difficulty with crafting culturally-neutral language.

The primary factor of our study was cultural matching, with two levels: *culturally-matched* (e.g., Sam used individualist script to interact with an individualist participant), and *culturally-mismatched* (e.g., Sam used individualist script to interact with a collectivist participant).

Table 1. Sam's full interaction script: common text is in the center, with the collectivist and individualist variants side by side.

	Collectivist	*Both*	*Individualist*
Intro		Hi there! Can we talk?	
		I am Sam. What is your name?	
		I am here to improve my algorithms and communication skills.	
Questions		Can I ask a few questions about you, first?	
		Are you currently a student?	
		Which country's culture do you identify closest to?	
		What is your favorite color?	
Backstory	Well, I am a member of the human-robot interaction team of our university.		Well, I am one of the first robots that is allowed to go around the university!
	Our team is one of the first teams to put a robot around our university. Have you ever heard of us?		I don't like to constantly ask for help from others in the lab. You know how that feels?
	Anyway, I am letting my team down by not knowing the answers to some questions, so I decided to improve my algorithms by volunteering for this experiment. Would you help me?		Anyway, I want to be independent from others, so I decided to improve my algorithms by volunteering for this experiment. Would you help me?
	Perfect! Your answers will help me make my team proud.		Perfect! Your answers will help me stand on my own feet.
Guide		Please answer the following questions by saying Yes, No, or Skip.	
Next Image Labelling Problem		Thanks! Here is the next one.	
	Thank you! Your answers help me be more useful in my team.		Thank you! Your answers help me be more independent.
		Sweet! And this one?	
	With your answers, our university team will get better results.		With your answers, I will be able to get better results on my own.
		Next question?	
	My team will be impressed!		I am impressed!
		What about this one?	
	You're on fire my friend! What about this one?		You're on fire dude! What about this one?
		You're the best! Next?	
	Your answers help me make my team proud!		Your answers help me stand on my own feet!
		You're on a roll!	
	Now I can support my teammates in the lab!		Now I can be on my own in the lab!
		Interesting! See if you can get this one? You seem to be skipping a lot of questions. Is anything wrong?	
Pleas	Can you please continue? We need more data.		Can you please continue? I need more data.
	But how do I make my team proud without your help? Please help me more!		But how do I become independent without your help? Please help me more!
Ending		Alright. Thank you for your help.	
		Please find the researcher at <location>, who will thank you with a 10$ gift card if you fill out a questionnaire about your experience.	
		Make Sure you have your completion code!	
		Have a good day!	

We used Sam's never-ending image labelling task to measure the impact of our manipulation. As this task is not likely rewarding for participants, we do not expect them to be intrinsically motivated to continue the task for long. We hypothesize that participants will relate better to Sam in the *culturally-matched* case and as such, will help it for longer, than in the *culturally-mismatched* case.

A. Procedure

We placed Sam in a public space at our local university, where it wandered around and approached passersby, attempting to engage interaction (Fig. 1). Sam kept a distance from the researcher and the camera (about 20 meters, Fig. 3), for a more ecologically valid interaction; we hoped people felt less monitored this way. We posted signs in the area, and pamphlets on Sam, to ensure people knew that a study was taking place and how to contact the researchers if they had questions.

Sam engaged and interacted with people as detailed in Section 3. Once a participant self-identified culture, Sam would match or mismatch randomly (near the end, Sam would try to balance the number of matched cases to the mismatched ones). We considered a person to be a participant if they engaged Sam long enough to answer at least one image-labelling question.

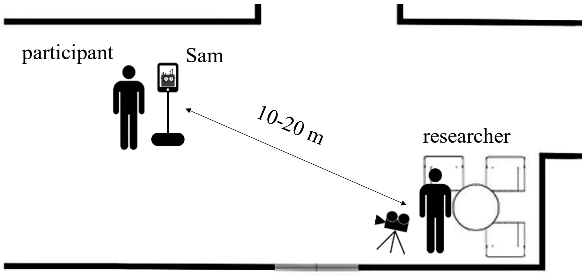

Fig. 3. The experiment space, with the researcher and camera at a distance from the robot engaging participants.

After interaction ended, Sam told the participant to find the researcher (to receive a gift card) and provided a short completion code (e.g., “M8EYP2”). Simultaneously, the on-site researcher approached the participant, and took them back to the camera location (Fig. 3) to avoid influencing other participants. The researcher administered the informed consent protocol and a series of post-test questionnaires, before debriefing on the study purpose and the robot (being remotely controlled). We asked participants to not share details with others, and we debriefed them in written form (participants could not keep the explanation), to avoid others hearing the details. This study was approved by the University of Manitoba Research Ethics Board.

B. Measures

Our primary measure was the interaction duration. That is, how long a person helped the robot. We measured from when the participant agreed to help to the instant that they stopped, measured from software logs. We did not use the number of image-labeling questions answered as during pilots we noted high individual difference in how long they spent scrutinizing an image before answering.

Post-test, we administered a demographic questionnaire including the person’s background, to be checked against what they told the robot for ensuring consistency. We also administered the Godspeed scale [7] to gain insight into participant’s perception of the robot.

C. Results

Sam interacted with 40 participants who answered at least one image-labelling question. 3 participants were excluded due to technical difficulties (e.g., network issue) and 2 indicated they knew the experiment purpose (i.e., heard from friend). 3 were outliers with times longer than 3IQR from the mean. This resulted in 32 participants (*M*=22.6 years, *SD*=4.4) from 12 countries (Table 2).

19 participants from Canada were classified as individualist, with the remaining 13 participants classified as collectivist (Table 2), resulting in 15 culturally-matched and 17 culturally-mismatched participants. Our post-test questionnaire did not indicate any errors in robot matching.

A *t*-test (one tailed) indicated a statistically-significant effect of culture matching on interaction duration, with culturally-matched participants helping Sam for longer (*M*=5m 34s, *SE*=2m 32s) than mismatched participants (*M*=4m 22s, *SE*=51s, $t_{17.6}$=-2.05, *ρ*=0.03, corrected as equal variances not assumed), indicating a 27% (1m 12s) increase and a medium to large effect size of *d*=.73 (Fig. 4).

Table 2. National cultures, number of participants. Only Canada, shaded, was individualist. Others were collectivist.

Canada, 19	Argentina, 1	Central America, 1
China, 2	Ethiopia, 1	India, 1
Iran, 1	Korea, 1	Nigeria, 2
Saudi Arabia, 1	Ukraine, 1	Vietnam, 1

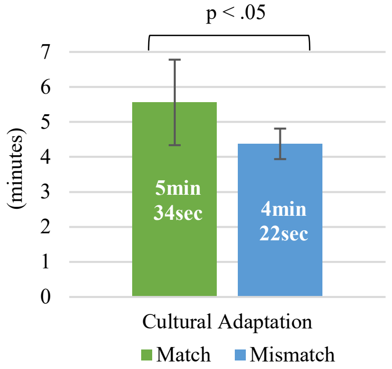

Fig. 5. Interaction duration for culturally matched vs mismatched participants. Error bars represent 95% CI.

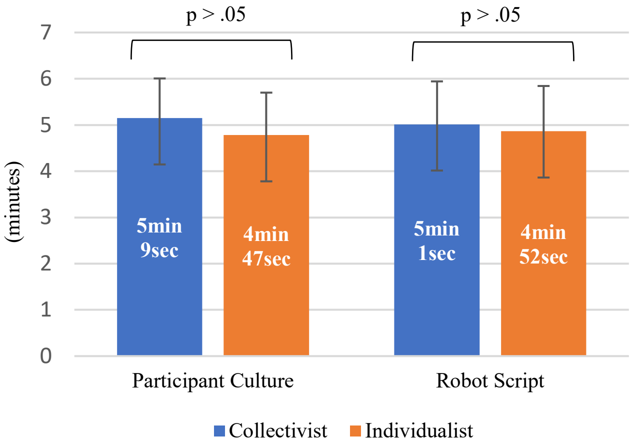

Fig. 4. Overall impact of participant culture and robot script on interaction time. Main effects not significant, interaction significant (p<.05). Error bars represent 95% CI.

Post-hoc, we investigated if overall participant culture or robot mode was a driving factor in the results, irrespective of match. A 2-way ANOVA (participant: individualist, collectivist, by Sam script: individualist, collectivist) found no main effect (*F*<1, Fig. 5). However, the interaction was statistically significant (*F*=5.09, *p*=.03); we address this in the discussion. We found no effect of matching, robot script, or participant culture, on their perceptions of the robot.

D. Discussion

Our results support our hypothesis that passersby will help a robot for longer (72s) when its script matches their cultural background, than when it mismatches. While our study does not uncover the mechanism behind this, it lends support to our strategy of getting people to relate to a robot by using language targeted to their cultural background. We suspect that this is simply the well established tendency for people to feel more socially connected to people similar to themselves (homophily, [cite]), which may have increased empathy for the robot’s situation, thus pressuring people to keep helping the robot.

However, we found no impact of matching (or participant culture, or robot script) on participant perceptions of the robot. Thus, even though participants did not reflect on the robot differently, they still interacted differently.

Our post-hoc analysis of the overall impact of either the robot script, or participant culture, aimed to uncover underlying drivers of our effect. Our results did not find effects of either participant culture or robot mood overall, suggesting that the result is not being driven by differences in the robot script or participant background independently.

From plotting the statistically-significant interaction between the two variables (Fig. 6) we can see that this interaction explains our main effect: the impact of the robot's script depends on the participant background culture. Put another way, it is the match or mismatch between participant culture and robot script that explains the result. However, we note that the interaction highlights the effect may be larger for the collectivist script than the individualist (Fig. 6), which requires further investigation.

Our results contribute to work of adaptive robots, and presents a new design technique that such robots can use. We believe it is important to emphasize how simple our method is. Without using advanced behavior models or artificial intelligence and only by changing a script based on roughly profiling the person's culture, our robot was able to get people to help it for significantly longer. This suggests that perhaps people may simply like to help a robot for longer when it matches their own experiences and expectations.

As such, we would ask the reader to reflect on this result: is it a success that our robot got people to help it for 27% longer, simply by changing a script? Or, is our robot inappropriately manipulative, given that the script changes were effectively lies (referred to goals and a group that did not exist), and there was no benefit to the human? We further examine this question at the end of the paper.

E. Limitations

Our interaction design and result were only initial iterations on a design strategy, and as such, there is a great deal of room for improvement. On the cultural side, our approach to matching participant background is quite simple, being based only on a single self-report question. Participants can easily misreport their background, and culture itself is more dynamic and nuanced than being reduced to a single binary classifier [57]. Work moving forward should develop more complex models that are able to adapt to cultures more flexibly and accurately.

Further, we only conducted the study in the host country. All people who identified with other cultures are exceptional: they are immigrants or visitors to Canada. Our work can be more accurately described as matching a collectivist script to collectivist people in an individualist society. Further inquiry should deploy robots embedded within target cultures to get a more generalizable result. As part of this, we only had fewer than twenty participants per condition; larger numbers in addition to a more diverse participant base is needed to reflect more strongly on the impact of cultural adaption.

Conducting our study in-the-wild enabled us to increase ecological validity, as participants were engaged naturally within their environment, without prior priming about the study. However, the flip side is that this introduced a great deal of noise due to the lack of ability to control interaction. Some people interacted individually, some in groups, and some had to leave prematurely due to other engagements. Some people wore headphones, some did not. We had more technical difficulties than would be expected in a lab. In addition, our study design still had the researcher and pamphlets on-site (see [cite your other workshop paper] for discussion), which unfortunately limits the ecological validity of the results.

V. Reflections

Our initial reaction to these results was quite positive: we created a robot that can adapt to people, using a coarse-grained flavor of cultural dimension theory and a very simple implementation, resulting in people helping it for 27% longer than if it did not match their background. However, as we watched our videos and analyzed our data, we increasingly started to notice a darker side to our results. In particular, some participants looked quite tired, and bored, and even then, hesitated to leave the interaction. Again, we note that there is very little benefit to the person in helping the robot beyond the novelty of seeing a robot in a public place; it is the robot that is getting more help. Thus, is this a positive use of social robotics, leveraging human-like interaction strategies to improve interaction? Or, is our robot's use of cultural theory manipulative, given that it is only for the robot's benefit, and it may be a detriment to participants?

To complicate the discussion, we note that our robot changes its backstory and motivators, just to match the person's background. For example, the collectivist script invokes a desire to help their team, yet there is no actual team in reality. This emulates the value of supporting in-groups and have the person relate to them. Our robot tells calculated lies in an attempt to get more help from the person, solely for its own ends.

We admit that this framing may be a little negative. Such interactions are common among people, such as a sales person feigning interest in a sports team to build comradery, or a professional leveraging small talk to build trust. More subtly, in a workplace a colleague may be uncharacteristically friendly and supportive, in preparation for asking a work favor. In our daily lives we accept a level of such behavior from people, so perhaps we will expect it of robots, too, and these behaviors may not be seen as manipulative. Further, as noted in Section 2, some robots are explicitly designed to be persuasive, for goals that the user may have subscribed to (such as saving energy [26]). In these cases, we respect people's choices to employ technology that can help them achieve a desired behavior or outcome.

This overarching question will become increasingly important as robotics advances, and social robots become more sophisticated and commonplace. The ability of such robots to deliberately employ manipulative social behavior will only increase. As such, we believe it is very important to develop and establish guidelines for how social robots should behave, and for helping us (robotics interface designers) decide when social interaction techniques are acceptable, and when they are not. Eventually, perhaps this will mirror what has happened with marketing and advertising, where regulations have been established regarding deceptive or unfair practices, for example, many locales have rules about subliminal priming [17] or targeting children [13]. We finish this paper by proposing a set of discussion points that we have landed on from our inquiry.

A. Discussion points

1) How to decide what is acceptable

As a community we should discuss what mechanisms we can use to decide which uses of social robots are acceptable, and which are not, and which criteria we can establish for determining when social techniques can be used.

In our own discussions, we often came back to the principle of who stands to benefit from a technique. That is, is a social technique used to benefit the users, such as helping them work effectively with a robot, understand its state, or be more comfortable. Or, is the technique used to benefit the robot or other stakeholders (e.g., by selling them something or getting work from them). How does the calculation change when it is a combination of benefitting both? That is, will people accept some manipulation (e.g., to sell them something) if they get a benefit (such as a robot carrying something for them), and how do we calculate this risk-benefit trade off?

Even when a technique is only used for the robot's benefit, some situations are more acceptable than others. For example, would a robot asking people for help because it is lost, and looking sad and forlorn to get attention, be considered devious?

Another angle is informed consent of users. Our robot presented in this paper clearly deceived people, both by lying about its purpose (trying to get more work from people, e.g. by stating that it was part of a group, etc.), and by hiding the cultural adaption and the reasons why it asked questions. This mirrors a recommendation that, with robots, their "machine nature" should be transparent to users to avoid exploiting them [18]. Perhaps a robot like ours, for example, should inform people that it will adapt to their culture in an attempt to be more relatable. Deception is also problematic in cases where it is used solely for a person's benefit. For example, perhaps a medical assistant robot can manipulate a person to encourage them to take medication that they do not want to take. Would this infringe on the patient's liberty?

2) Is it appropriate for robots to act like people?

In daily life as part of normal social interaction we regularly use behaviors directly designed to shape others' moods, reactions, and outcomes of interaction. This ranges from long-term alliance strategizing at work, to doing favors at home to soften the blow of a mistake. This is often expected of people, and very often such strategies back-fire and someone can be labelled manipulative or deceptive. However, this happens within the complex context of human relationships.

Just because people use social strategies as such, does this mean that it is okay for a robot to act similarly? Given that a robot is a cold calculating machine (in contrast to a human with their own empathy and emotional system), is it ever appropriate for a robot to attempt to shape outcomes using social strategies? This issue has been raised under the umbrella of "inauthentic" interactions [56,63,65], which highlights the inherent mismatch with a non-human entity (a robot) using human-like techniques. This leads to incongruent interactions, such as a robot presenting itself as having an emotional system, but not reacting in appropriately when it is threatened. This can lead to unnatural and perhaps unhealthy relationships with the robot (e.g., mimicking abuse). Finally, unlike humans, robots have excellent memory and can apply algorithms perfectly. For example, a shopping-mall robot may have access to a person's long-term shopping history, can read their facial expressions, notice who they are with, and make a calculated social gesture meant to sell the person a product. One could imagine a similar robot in a casino. Thus, the potential capabilities of robots in regard to manipulation seems to outstrip what a human could possibly do, making the comparison problematic.

3) How do social robots differ from existing intelligent technologies and marketing strategies?

In both traditional media and more modern online information spaces, we are constantly inundated with targeted marketing, smart bots, and long-term campaigns designed to earn our dollars, shape our opinions, or our behaviors. Do our arguments simply fall under this larger, contemporary discussion about limits and dangers of technology? Or, do social robots, with their active physical presence and embodiment within our personal spaces, and tendencies surrounding anthropomorphism [71], require special attention?

VI. CONCLUSION

We designed, implemented, and deployed Sam, an adaptive robot that changes its script based on a user's cultural background in an attempt to get the user to help it for longer. Sam was successful, with participants on-average helping Sam for 72 seconds longer when it matched their culture (e.g., Sam using an individualist script with an individualist participant), than when it mismatched.

While initially this work served as a simple proof of concept for robots leveraging minimal information about a user's background to improve interaction, through reflecting on our own work we started to see the results differently. We developed a worry that perhaps our work instead is a demonstration of just how easily robots can use social interaction techniques to deceive and manipulate people into helping them for longer. This project fed a lively discussion and debate among our researchers about whether the robot was simply good interaction design, or, Machiavellian, being deceitful and manipulative for its own ends.

As such we use this paper as an opportunity to reflect on this question. In addition to detailing our core study design, methodology, and results, we pose discussion points that emerged from conducting this study. We envision that this can be a catalyst to generate discussion within the Human-Robot Interaction community.

REFERENCES

[1] Iina Aaltonen, Anne Arvola, Päivi Heikkilä, and Hanna Lammi. 2017. Hello Pepper, May I Tickle You? Children's and Adults' Responses to an Entertainment Robot at a Shopping Mall. *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, March 6-9: 53–54.

[2] Abhijeet Agnihotri, Alison Shutterly, Abrar Fallatah, Brian Layng, and Heather Knight. 2018. ChairBot Café : Personality-Based Expressive Motion. *In adjunct Proceedings of the 13th Annual ACM/IEEE International Conference on Human-Robot Interaction . ACM.*

[3] Sean Andrist, Micheline Ziadee, Halim Boukaram, Bilge Mutlu, and Majd Sakr. 2015. Effects of Culture on the Credibility of Robot Speech.

Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '15: 157–164.

[4] John A. Bargh, Mark Chen, and Lara Burrows. 1996. Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology* 71, 2: 230–244.

[5] Christoph Bartneck, Timo Bleeker, Jeroen Bun, Pepijn Fens, and Lynyrd Riet. 2010. The influence of robot anthropomorphism on the feelings of embarrassment when interacting with robots. *Paladyn, Journal of Behavioral Robotics* 1, 2: 109–115.

[6] Christoph Bartneck, Michel van der Hoek, Omar Mubin, and Abdullah Al Mahmud. 2007. "Daisy, Daisy, give me your answer do!" *Proceeding of the ACM/IEEE international conference on Human-robot interaction - HRI '07*: 217.

[7] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics* 1, 1: 71–81.

[8] Christoph Bartneck, Tatsuya Nomura, Takayuki Kanda, Tomohiro Suzuki, and Kennsuke Kato. 2005. Cultural differences in attitudes towards robots. *AISB'05 Convention: Social Intelligence and Interaction in Animals, Robots and Agents - Proceedings of the Symposium on Robot Companions: Hard Problems and Open Challenges in Robot-Human Interaction*, February 2016: 1–4.

[9] Paul Baxter, James Kennedy, Emmanuel Senft, Séverin Lemaignan, and Tony Belpaeme. 2016. From characterising three years of HRI to methodology and reporting recommendations. *ACM/IEEE International Conference on Human-Robot Interaction* 2016–April: 391–398.

[10] Beatrice Blyth Whiting. 1980. Culture and social behavior: a model for the development of social behavior.

[11] Bastiaan Van den Berg. 2012. Differences between Germans and Dutch People in Perception of Social Robots and the Tasks Robots Perform. *16th Twente Student Conference on IT*: 1–6.

[12] Serena Booth, James Tompkin, Hanspeter Pfister, Jim Waldo, Krzysztof Gajos, and Radhika Nagpal. 2017. Piggybacking robots: human-robot overtrust in university dormitory security. 426–434.

[13] Almudena Gonzalez Dell Valle. 2013. A reflection on European regulation of television advertising to children. *Communication Research Trends*.

[14] Daniel R. Denison and Gretchen M. Spreitzer. 1991. Organizational culture and organizational development: A competing values approach. *Research in Organizational Change and Development 5*, 1–21.

[15] Tamara Denning, Cynthia Matuszek, Karl Koscher, Joshua R. Smith, and Tadayoshi Kohno. 2009. A spotlight on security and privacy risks with future household robots. *Proceedings of the 11th international conference on Ubiquitous computing - Ubicomp '09*: 105.

[16] Amol Deshmukh, Sooraj Krishna, Nagarajan Akshay, Vennila Vilvanathan, Sivaprasad J V, and Rao R Bhavani. 2018. "In the wild" - In Rural India. *In adjunct Proceedings of the 13th Annual ACM/IEEE International Conference on Human-Robot Interaction . ACM.*, 5.

[17] Stephanie Dube. Laws on Subliminal Marketing.

[18] EPSRC. 2010. Principles of robotics.

[19] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. 2003. A survey of socially interactive robots. *Robotics and Autonomous Systems* 42, 3–4: 143–166.

[20] Joel Garreau. 2007. Bots on the Ground. *Washington Post, WWW, \url{http://www.washingtonpost.com/wp-dyn/content/article/2007/05/05/AR2007050501009_pf.html}, Visited April 9th, 2008.*

[21] Denise Y Geiskkovitch, Derek Cormier, Stela H Seo, and James E Young. 2016. Please Continue, We Need More Data: An Exploration of Obedience to Robots. *Journal of Human-Robot Interaction* 5, 1: 82–99.

[22] Saul Greenberg, Sebastian Boring, Jo Vermeulen, and Jakub Dostal. 2014. Dark patterns in proxemic interactions: A critical perspective. *2014 Conference on Designing Interactive Systems (DIS'14)*: 523–532.

[23] and Elizabeth Chua Gudykunst, William B., Stella Ting-Toomey. 1988. *Culture and interpersonal communication*. Sage Publications.

[24] William B Gudykunst and Stella Ting-toomey. 1996. The Influence of Cultural and Individual Values on Communication Styles Across Cultures. 22, 4.

[25] Edward Twitchell Hall. 1989. *Beyond culture*. Anchor.

[26] Jaap Ham and Cees Midden. 2010. A persuasive robotic agent to save energy: The influence of social feedback, feedback valence and task similarity on energy conservation behavior. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 6414 LNAI: 335–344.

[27] Kerstin Sophie Haring, David Silvera-Tawil, Yoshio Matsumoto, Mari Velonaki, and Katsumi Watanabe. 2014. Perception of an android robot in Japan and Australia: A cross-cultural comparison. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8755: 166–175.

[28] Frank Hegel, Claudia Muhl, Britta Wrede, Martina Hielscher-fastabend, and Gerhard Sagerer. 2009. Understanding Social Robots. Section II.

[29] M. Hofstede, G., Hofstede, G. J. & Minkov. 2010. *Cultures and Organizations: Software of the Mind (Rev. 3rd ed.)*. New York: McGraw-Hill.

[30] Geert Hofstede. 2001. Culture's Consequences: Comparing Values, Behaviors, Institutions and Organisations Across Nations. 27, 1: 89–94.

[31] Geert Hofstede. 2011. Dimensionalizing Cultures : The Hofstede Model in Context Dimensionalizing Cultures : The Hofstede Model in Context. 2: 1–26.

[32] Li Jun Ji, Zhiyong Zhang, and Richard Nisbett. 2004. Is it culture or is it language? Examination of language effects in cross-cultural research on categorization. *Journal of Personality and Social Psychology* 87,1:57-65.

[33] Malte F. Jung, Nikolas Martelaro, and Pamela J. Hinds. 2015. Using Robots to Moderate Team Conflict. In *Proc. ACM/IEEE international conference on Human-Robot Interaction - HRI '15*, 229–236.

[34] Malte Jung and Pamela Hinds. 2018. Robots in the Wild: A Time for More Robust Theories of Human-Robot Interaction. *ACM Transactions on Human-Robot Interaction (THRI)* 7, 1: 2.

[35] Hee Rin Lee and Selma Sabanović. 2014. Culturally variable preferences for robot design and use in South Korea, Turkey, and the United States. *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction - HRI '14*: 17–24.

[36] Hee Rin Lee, Jayoung Sung, Selma Sabanovic, and Joenghye Han. 2012. Cultural design of domestic robots: A study of user expectations in Korea and the United States. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*: 803–808.

[37] Namyeon Lee, Jeonghun Kim, Eunji Kim, and Ohbyung Kwon. 2017. The Influence of Politeness Behavior on User Compliance with Social Robots in a Healthcare Service Setting. *International Journal of Social Robotics* 9, 5: 727–743.

[38] Rosemarijn Looije, Mark A. Neerincx, and Fokie Cnossen. 2010. Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors. *International Journal of Human Computer Studies* 68, 6: 386–397.

[39] David Matsumoto and Seung Hee Yoo. 2006. Toward a New Generation of Cross-Cultural. *Perspectives on Psychological Science* 1, 3: 234–250.

[40] Miller McPherson, Lynn Smith-Lovin, and James M Cook. 2001. Birds of a Feather : Homophily in Social Networks http://www.jstor.org/stab. *Annual Review of Sociology* 27, 2001: 415–444.

[41] Cees Midden and Jaap Ham. 2009. Using negative and positive social feedback from a robotic agent to save energy. *Proceedings of the 4th International Conference on Persuasive Technology - Persuasive '09*: 1.

[42] Tatsuya Nomura. 2014. Comparison on negative attitude toward robots and related factors between Japan and the UK. 87–90.

[43] Tatsuya Nomura. 2015. General Republics' Opinions on Robot Ethics: Comparison between Japan, the USA, Germany, and France. In *4 th International Symposium on New Frontiers in Human-Robot Interaction*, 12–16.

[44] Tatsuya Nomura. 2017. Cultural Differences in Social Acceptance of Robots *.

[45] Tatsuya Nomura, Tomohiro Suzuki, Takayuki Kanda, Jeonghye Han, Namin Shin, Jennifer Burke, and Kensuke Kato. 2008. What People Assume About Humanoid and Animal-Type Robots: Cross-Cultural

Analysis Between Japan, Korea, and the United States. *International Journal of Humanoid Robotics* 05, 01: 25–46.

[46] Tatsuya Nomura, Dag Sverre Syrdal, and Kerstin Dautenhahn. 2015. Differences on Social Acceptance of Humanoid Robots between Japan and the UK. *4th International Symposium on New Frontiers in Human-Robot Interaction*.

[47] Brittany Postnikoff and Ian Goldberg. 2018. Robot Social Engineering: Attacking Human Factors with Non-Human Actors.

[48] A Putten, N Kramer, L Hoffmann, S Sobieraj, and S Eimler. 2013. An Experimental Study on Emotional Reactions Towards a Robot. *International Journal of Social Robotics* 5, 1: 17–34.

[49] P. L Patrick Rau, Ye Li, and Dingjun Li. 2009. Effects of communication style and culture on ability to accept recommendations from robots. *Computers in Human Behavior* 25, 2: 587–595.

[50] P. L Patrick Rau, Ye Li, and Dingjun Li. 2010. A cross-cultural study: Effect of robot appearance and task. *International Journal of Social Robotics* 2, 2: 175–186.

[51] Laurel Riek. 2012. Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines. *Journal of Human-Robot Interaction* 1, 1: 119–136.

[52] Selma Sabanovic, Marek P. Michalowski, and Reid Simmons. 2006. Robots in the wild: Observing human-robot social interaction outside the lab. *International Workshop on Advanced Motion Control, AMC* 2006: 576–581.

[53] Daisuke Sakamoto and Tetsuo Ono. 2006. Sociality of Robots: Do Robots Construct or Collapse Human Relations? In *Proc. ACM/IEEE international conference on Human-Robot Interaction - HRI '06*, 355–356.

[54] Elaheh Sanoubari and James E. Young. 2018. Hi human, can we talk? An in-the-wild study template for robots approaching unsuspecting participants. In *Workshop on the Social Robots in the Wild at the International Conference on Human-Robot Interaction*.

[55] Stela H. Seo, Denise Geiskkovitch, Masayuki Nakane, Corey King, and James E. Young. 2015. Poor Thing! Would You Feel Sorry for a Simulated Robot? In *Proc. International Conference on Human-Robot Interaction - HRI '15*, 125–132.

[56] Amanda Sharkey and Noel Sharkey. 2012. Granny and the robots: Ethical issues in robot care for the elderly. *Ethics and Information Technology* 14, 1: 27–40.

[57] Shennan. 2000. Population, Culture History, and the Dynamics of Culture Change. *Current Anthropology* 41, 5: 811.

[58] Takanori Shibata and Kazuyoshi Wada. 2011. Robot therapy: A new approach for mental healthcare of the elderly - A mini-review. *Gerontology* 57, 4: 378–386.

[59] Reid Simmons, Maxim Makatchev, Rachel Kirby, Min Kyung Lee, Imran Fanaswala, Brett Browning, Jodi Forlizzi, and Majd Sakr. 2011. Believable Robot Characters. *AI Magazine* 32, 4: 39.

[60] Smith, P. B., Trompenaars, F., and S. Dugan. 1995. The Rotter locus of control scale in 43 countries: A test of cultural relativity. *International Journal of Psychology, 30(3), 377-400.*

[61] Yutaka Suzuki, Lisa Galli, Ayaka Ikeda, Shoji Itakura, and Michiteru Kitazaki. 2015. Measuring empathy for human and robot hand pain using electroencephalography. *Scientific Reports* 5: 1–9.

[62] Haodan Tan, Liping Sun, and Selma Sabanovic. 2016. Feeling green: Empathy affects perceptions of usefulness and intention to use a robotic recycling bin. *25th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2016*: 1051–1056.

[63] Sherry Turkle. 2017. Why these friendly robots can't be good friends to our kids.

[64] Sherry Turkle, Cynthia Breazeal, Olivia Dasté, and Brian Scassellati. 2006. Encounters with kismet and cog: Children respond to relational artifacts. *Digital Media: Transformations in Human Communication*, September: 1–20.

[65] Sherry Turkle, Will Taggart, Cory D. Kidd, and Olivia Dasté. 2006. Relational artifacts with children and elders: The complexities of cybercompanionship. *Connection Science* 18, 4: 347–361.

[66] Gentiane Venture, Bipin Indurkhya, and Takamune Izui. 2017. Dance with Me! Child-Robot Interaction in the Wild. *International Conference on Social Robotics* 10652: 375–382.

[67] Lin Wang, Pei-Luen Patrick Rau, Vanessa Evers, Benjamin Krisper Robinson, and Pamela Hinds. 2010. When in Rome: The role of culture & context in adherence to robot recommendations. *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*: 359–366.

[68] Astrid Weiss, Betsy Van Dijk, and Vanessa Evers. 2012. Knowing me knowing you: Exploring effects of culture and context on perception of robot personality. *International Conference for the Information Community 2012*: 133–136.

[69] Ehren Wolfe, Jerry Weinberg, and Stephen Hupp. 2018. Deploying a Social Robot to Co-Teach Social Emotional Learning in the Early Childhood Classroom. *In adjunct Proceedings of the 13th Annual ACM/IEEE International Conference on Human-Robot Interaction . ACM.*

[70] James E. Young, Richard Hawkins, Ehud Sharlin, and Takeo Igarashi. 2009. Toward acceptable domestic robots: Applying insights from social psychology. *International Journal of Social Robotics* 1, 1: 95–108.

[71] James E Young, Jayoung Sung, Amy Voida, Ehud Sharlin, Takeo Igarashi, Henrik I Christensen, and Rebecca E Grinter. 2010. Evaluating Human-Robot Interaction. *International Journal of Social Robotics* 3, 1: 53–67.

[72] "Culture." In Merriam-Webster's dictionary.