Are Robots Ready for Administering Health Status Surveys? First Results from an HRI Study with Subjects with Parkinson's Disease

Priscilla Briggs
HRI Laboratory
Dept. of Computer Science
Tufts University
161 College Avenue
Medford, MA, USA
priscilla.briggs@tufts.edu

Matthias Scheutz
HRI Laboratory
Dept. of Computer Science
Tufts University
161 College Avenue
Medford, MA, USA
matthias.scheutz@tufts.edu

Linda Tickle-Degnen
Dept. of Occupational Therapy
Tufts University
26 Winthrop Street
Medford, MA, USA
linda.tickle_degnen@tufts.edu

ABSTRACT

Facial masking is a symptom of Parkinson's disease (PD) in which humans lose the ability to quickly create refined facial expressions. This difficulty of people with PD can be mistaken for apathy or dishonesty by their caregivers and lead to a breakdown in social relationships. We envision future "robot mediators" that could ease tensions in these caregiver-client relationships by intervening when interactions go awry. However, it is currently unknown whether people with PD would even accept a robot as part of their healthcare processes. We thus conducted a first human-robot interaction study to assess the extent to which people with PD are willing to discuss their health status with a robot. We specifically compared a robot interviewer to a human interviewer in a within-subjects design that allowed us to control for individual differences of the subjects with PD caused by their individual disease progression. We found that participants overall reacted positively to the robot, even though they preferred interactions with the human interviewer. Importantly, the robot performed at a human level at maintaining the participants' dignity, which is critical for future social mediator robots for people with PD.

Categories and Subject Descriptors

H.4m [**Information Systems Applications**]: Miscellaneous

Keywords

Assistive robots; Parkinson's disease; quantitative field study

1. INTRODUCTION

Parkinson's Disease (PD) is a degenerative disability which affects 1.6% of people over 65 in the United States [22]. PD is characterized by a decline in motor control and for some patients this decline can lead to "facial masking", a condition in which nuanced control of facial expression is partially or completely extinguished [17]. People exhibiting facial masking are perceived by both laypeople and healthcare providers alike as having less desirable personalities and as being less competent [19]. The lack of facial and other affective signals can be, wrongly, processed by the caregiver as lack of interest, affect, and engagement, and potentially lack of truth on the side of the person with PD [16]. As a result, communicative interactions between people with PD and their caregivers can go awry, leading to feelings of *stigmatization* and, ultimately, a loss of social control [18], causing a degradation of self dignity. An important goal for the management of PD is thus to break this "loop of miscommunication" to maintain the dignity of the person with PD.

One way to achieve these goals might be to use robots that can facilitate the interactions in the client-caregiver relationship by intervening when facial masking created miscommunication. The robot would thus function as a "mediator" (rather than a replacement of the caregiver) with the sole purpose of ensuring a functioning relationship between a person with PD and the caregiver. This, in turn, would reduce the risk for stigmatization.

A prerequisite for developing robots that could serve such a mediator role for PD therapy in the future, however, is knowing whether people with PD would be open to a "working alliance" with a robot in the first place. A *working alliance* has been defined in the context of therapy as a relationship in which both the client and the practitioner are "working together to achieve agreed upon goals through agreed upon methods of interventions" [15]. Hence, before we can start the development of a mediator robot, or any robot that can support people with PD for that matter, we need to find out whether people with PD are willing to pursue the goal of managing their disease together with a robot. One such way is to interact with a robot that can ask them questions about their health in natural language and is co-present when they talk with their caregivers. Hence, the goal of the present study was to perform the first empirical investigation of this kind, i.e., exploring whether people with PD would be open to robot mediators that ask them health-related questions.

The rest of the paper is structured as follows. We start with a brief description of the problems involved in PD and

HRI'15 March 02–05, 2015, Portland, OR, USA
Copyright 2015 ACM 978-1-4503-2883-8/15/03 ...$15.00.
http://dx.doi.org/10.1145/2696454.2696476.

the challenges posed by PD for robotics, including a summary of related work in HRI. We then introduce the experimental setting and describe the interaction used in this study which allowed us to analyze the reactions of people with PD to the robot interviewer. Next we present the methods used to analyze the data, followed by the observed results. Finally, we discuss the implications of our findings for future studies and robotic developments in the domain of PD.

2. BACKGROUND

One of the principle aims of health care delivery is to optimize participation in preferred social roles [21]. People who do not engage in sufficient social activity are at risk for unfavorable health outcomes such as accelerated motor decline [3]. This problem led to the idea of *social self-management of health*, which is considered to be any self-care practices towards the goal of social comfort while simultaneously maintaining both physical and mental well-being [17]. Social self-management health practices include maintaining relationships and seeking help when it is needed [18]. A robot may be able to improve the social self-management of a person with PD. Yet, there is no published research to date that shows whether and how a robot could help people with PD to manage their social self, in particular, people with PD who exhibit facial masking.

Facial masking is a threat to the social self-management of a person with PD that can add mistrust into a relationship, or cause others to believe people with PD are less cognitively capable than they actually are [19]. If people could be trained to ignore the misleading facial expressions and pay attention instead to what the person with PD is saying, which more accurately represents what their state of mind and their feelings, then relationships could avoid being spoiled by facial masking. Unfortunately, even trained professionals are unconsciously and incorrectly reliant on automatic reactions to a masked face, even when they consciously know that they should not [19]. A robot, on the other hand, would have no problems discounting the expressions of a person exhibiting facial masking and could restrict itself to only analyzing the content of a person's speech interactions. It is thus possible that robots could eventually create a more accurate model of the emotional state of the person with PD than a layperson or even a medical professional could. If the robot were then able to communicate that person's emotions to the caretaker, the robot might be able to alleviate some of the stigmatization caused by facial masking in the caregiver-client interaction.

Our overarching goal is to work towards creating such a robot that can maintain the dignity and autonomy of people with PD, mainly by alleviating the effects of facial masking. As a first step, we designed the experiment described in this paper to assesses the ability and utility of a robot to interact with a person with PD. We aimed to evaluate whether a robot would induce feelings of loss of control or promote feelings of competency. We also wished to find out whether people with PD would like to interact with a robot and would feel that it acted in an attentive and understanding manner, all of which are prerequisite for forming a working alliance between a human with PD and a robot.

3. RELATED WORK

There are some indications that people with PD might be open to interacting with robots as recent research in HRI suggests that older adults without PD are generally interested in having robots assist them with their health-care. A survey of people living or working in a retirement village, for example, showed that older people were positive about the possible future role of robots in the village as long as they were easy to use, addressed a need, and assisted in independence and dignity [2]. When introduced to a robot that could take medical measurements, encourage physical activity, and discuss environmental hazards, participants felt highly positively towards it and were excited about the possibilities going forwards [20].

Given the potential of and interest in robots designed to facilitate physical abilities in older adults and people with disabilities, there is a growing body of research investigating how robots might be able to help. For example, robot coaches have been created that can motivate older adults to exercise [7]. Robotic canes can provide greater mobility assistance than normal walkers for people who wish to stay in their homes longer [6]. Moreover, work with a robot practicing tasks set by a physical therapist for stroke patients has shown increases in the amount of mobility for participants [10]. And there has even been some work in providing new technologies to people with PD such as using Google Glass as an assistive device [12].

Much less work, however, has investigated how to design robots to interact *socially* with those with disabilities. Most research in this area focuses on children with Autism Spectrum Disorder (ASD). Children with ASD interacting with a social robot, for example, exhibited more social behavior (in both human-human and human-robot interactions) than when they were with a randomly moving (non-social) robot [8]. Interacting with a social robot also elicited more social behaviors than interacting with a human adult or a tablet game [11].

These examples are among several encouraging findings that social robots can increase the amount of social activity in children with ASD and one could be tempted to apply robots developed for people with ASD to people with PD. However, ASD and PD are very different conditions and success for children with ASD does not guarantee the same success with people with PD. For example, ASD is characterized by a lack of simple social skills and an aversion against interacting with other people. In contrast, PD is primarily a physical, not cognitive disability, and people with PD generally are socially competent. We are thus in the early stages of investigating whether robots can play a role in mediating social human-human interaction for a group of people who cannot express their emotions adequately through facial expressions or affect in their voice (e.g., see [1] for our first preliminary report on human-robot interaction between robots and people with PD).

4. METHODS

In this study we aimed to test whether people with PD would accept a robot in a healthcare function and to show this acceptance in both participant self report and behavior. We chose a task for the robot that is very familiar to people with PD: assessing their disease-related quality of life using the Parkinson's Disease Questionnaire-39 (PDQ-39), a standardized measurement tool used by healthcare professionals.

We designed a within-subjects experiment in which half the questions of the PDQ-39 were administered by a robot and half of the questions by a human. This allowed us to compare the acceptability of the robot to a human and thus to control for individual differences in the subjects. The human interviewer for all participants was a female graduate student in Occupational Therapy trained in administering health interviews.

Note that the within-subjects design is critical for the assessment of behavior during interviews given the individual variability of the motor problems of people with PD. By having the robot and the human both administer half the PDQ-39, we were able to look at measures such as eye gaze and speech within-subjects.

It is also important to stress that experiments involving participants with PD cannot be compared to experiments with typical undergraduates. Not only are there significant differences in age, life experience and disease status, but also in the reason for doing the experiment and the degree to which the population is invested in the task. If the task is directly related to a debilitating disease experienced by the participant, then the participant will be more invested in the task and have a higher threshold for participating in the research in the first place compared to a healthy undergraduate student. Because of the investment that people with PD have in their healthcare, it is thus more difficult to recruit people with PD for such experiments where a robot takes on an important role in their personal health management process.

4.1 Participants

We were able to recruit 17 participants with PD (3 women, 1 unknown[1]) from PD support groups and fundraiser events, and from talks given at older adult community services. To be included in the study, participants had to self-report having PD, be able to understand and answer questions about their quality of life, be able to travel to the research study location, and be willing and able to provide informed consent. All participants were in the early to moderate stages of PD and were completely cognitively intact.

The robot malfunctioned during four of the 17 interviews. Data from these interviews were not included in any analyses reported here. Thus, the data from a total of 13 participants (one woman, one unknown) with a robot functioning according to protocol were analyzed for the current study.

The majority of participants were male, which is typical in studies of people with PD. The average age of participants was 67.77 years old ($SD = 14.13$). Of the 12 participants who reported their education, 8 had a college degree. The average participant had been diagnosed with PD 6.64 years ago ($SD = 5.36$, range = 5months - 17 years). No participants reported having interacted with a robot before.

Though this sample is small, it should be noted that it is a time-intensive operation to recruit subjects from a target population such as people with PD. Not only is there a smaller population of people with PD than many other populations, but scheduling becomes more difficult as participants often have difficulty with mobility. This study involves a greater effort level and potential payoff compared to other studies in HRI. First, it applies HRI with real robots instead of on Mechanical Turk. Secondly, it is done with a clinical population instead of healthy subjects at the university such as undergraduate students. Hence, it is not surprising (even though less than ideal) that the subject pool turned out to be fairly homogeneous and not gender-balanced (as is otherwise fairly straightforward to ensure).

[1]All participants were given the option to choose from the genders of "Male", "Female", "Unknown" or "Unspecified" on their demographic survey. One participant is listed as having an unknown gender because that participant chose the "Unknown" option.

Figure 1: The robot interviewer was the Nao robot by Aldebaran.

4.2 Materials

The Nao robot by Aldebaran (Figure 1) was used as the robot interviewer. Its small size and expressive capabilities are typically seen as likable and easy to engage with. A researcher teleoperated the robot out of sight of the participants, using the Aldebaran *Choreographe* module. As seen in Figure 2, Choreographe lets users design and run robot behaviors. When the play button on a box is selected, that box's effect will take place, followed by the effects of the boxes connected to it. Figure 2 shows a simple example. After the robot finishes asking PDQ-39 question number nine (box "Q9"), the program proceeds to wait (the "wait" box) as the participant answers the question. When the participant finishes their response, the experimenter can tell Choregraphe to stop wating and move onto the next box, which in this case is the "Alright" box, which causes the robot to say "Alright" in response to the participant's answer to the question. Then the robot moves on to asking question eleven (box "Q11") while making a hand gesture ("hand up & open" box). When the hand gesture is done and the question has been asked ("Wait Till Done" box), the robot goes back to its standard pose ("Standard Pose" box). Also visable in Figure 2 are some of the other response options that the experimenter can make the robot say, such as "OK", and "Wonderful". Additional idle movements, such as blinking, that were not related to the question at hand, happened at random times.

In addition to hand gestures and idle movements, the robot showed its mobility at the beginning of the interview by standing up while waving and introducing itself, then taking a few steps walking towards the participant and eventually sitting down to conduct the interview. At the end of

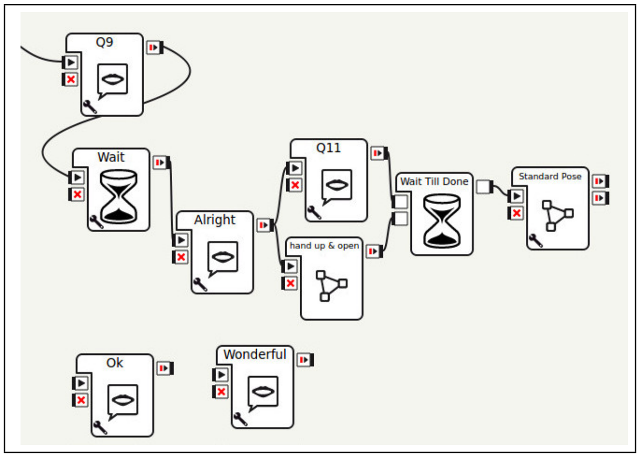

Figure 2: The Choreographe Module used to control the robot.

the task, the robot waved at the participant while saying "goodbye".

4.3 Procedure

The participant gave informed consent through procedures authorized by Tufts University Social, Behavioral and Educational Institutional Review Board. Following consent, the interviews began. Whether the participant was interviewed first by the robot or first by the human was a factor counterbalanced between participants. Interviews were video recorded. In order for the human and robot to have the same dialogue capabilities, we restricted the human interviewer to saying only phrases which the robot was programmed to be able to say. These included the interview questions, standardized introductory and ending remarks and 48 phrases such as "You said 'Always', is that correct?", "That sounds difficult", and "Can you repeat that?"[2]. The human interviewer held a copy of the interview questions, introductions, ending remarks and 48 phrases during the interview, as in a normal PDQ-39 interview, in which she would have had a copy of the questions. The human interviewer was instructed to use the same body language or gestures that she would usually use during this type of interaction.

Each interview consisted of the administration of a modified version of the PDQ-39, a measure of subjective health status and quality of life [9] commonly used in health practice and research with PD. It assesses eight domains of health and well-being important to people living with PD (Table 1) and is easy to administer by either human or robot. We split the PDQ-39 into two equal sets of questions from each of the eight domains of concern. Participants answered the degree to which they experienced the effect of PD on their lives by choosing one of five options along a Likert scale: "Always", "Often", "Sometimes", "Rarely" or "Never". One question about social support was present in both interviews in order to test if participants would answer it identically to both interviewers. This made a total of 20 items per half. After administration of the PDQ-39, the interviewer asked "Can you describe one of your favorite activities to me?", and would listen, only making small comments such as "That's wonderful". If the participant had an answer that was only a few words long the interviewer could say "Can you elaborate?" or "Can you describe that to me more?". This question allowed us to assess the content and quality of the participant's open-ended speech with the robot relative to the human.

A typical interaction with the either the robot or human interviewer could have included the following dialogue: The interviewer asks, "Due to having Parkinson's disease, how often during the past 30 days have you felt frightened or worried about falling in public?" The participant would then respond with one of the Likert scale items such as, "Sometimes." The interviewer would acknowledge this by saying, "OK" and then move onto the next question saying, "Due to having Parkinson's disease, how often during the past 30 days have you had difficulty showering and bathing?" If the participant said "Well, now that I live in an assisted living facility, I get help every day" the interviewer is free to ask the participant to choose a Likert option by saying, "Would you consider that to be, Always, Often, Sometimes, Rarely, or Never?" The participant might say, "Oh, I think I would consider that never." The interviewer would then acknowledge this with another "OK" and would move on with the interview.

After each interview, participants responded to written questionnaires asking about their experience and their perceptions of the interviewer. Participants filled out a demographic questionnaire (age, gender, education, duration of PD, and whether or not they had ever interacted with a robot before). Participants were compensated $25. Lastly, participants were debriefed about why the study was done and were given the opportunity to ask questions about the study.

Category	**Question** (Each question begins with "Due to having Parkinson's disease, how often during the past 30 days have you...")
Mobility	...had difficulty getting around in public places?
ADL	...had difficulty holding a drink without spilling it?
Emotions	...felt depressed?
Stigma	...felt embarrassed in public?
Social	...received the support you needed from your family or close friends?
Cognitions	...had distressing dreams or hallucinations?
Communication	...had difficulty speaking?
Body Pain	...had painful muscle cramps or spasms?

Table 1: A few of the questions asked by the interviewers in this study. All questions were from a modified version of the PDQ-39, which measures health quality of life for people with PD in the eight dimensions shown in the table.

4.4 Questionnaire Measures

Participants completed a series of written questionnaires after each interview, encompassing 41 items divided into four

[2]Although the human interviewer was for the most part successful in limiting herself to the required phrases, for some participants she made negligible, single-word errors. One example would be when if she asked a participant what his hobbies were he said "Reading", and she repeated back "Reading". It seems reasonable to assume that these small deviations of script did not significantly affect perceptions of the human interviewer.

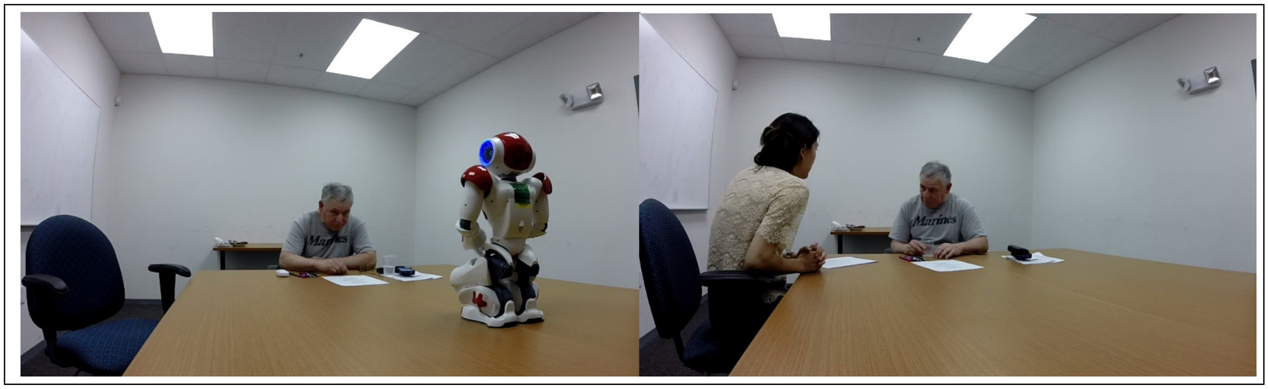

Figure 3: For this interaction, both a robot and a human administered a short questionnaire to a person with Parkinson's disease.

categories derived from previous research on the quality of the experience of interpersonal interaction between humans [13, 14]. Five-point Likert questions from 1 (not at all) to 5 (very), assessed the degree to which the participant, during each interview, experienced positive mood, optimal experience, an emotional bond to the interviewer, and perceived the interviewer as producing actions that create rapport. The internal consistency of items representing each category was assessed using Cronbach's alpha. Composite scores for each category were created by averaging all items after reversing the items noted below. The four categories and the items that comprise them are the following:

- **Positive Mood**. These items assessed the degree to which the participant felt alert, happy, irritable (reversed), involved, tense (reversed), sociable, bored (reversed), confused (reversed), cooperative, and embarrassed (reversed). The items demonstrated high internal consistency (Robot = .90; Human = .78).

- **Optimal Experience**. These items assessed the degree to which the participant experienced one's skills as having been fully engaged with competency and satisfaction during the interview [5]. Items were "I was able to concentrate", "It was hard to concentrate" (reversed), "I felt self-conscious" (reversed), "'I was aware of the camera" (reversed), "The interview was challenging to me" (reversed), "I felt skillful during the interview", "I felt in control of the situation", "I did a successful interview", and "I felt good about myself". The items demonstrated moderate to high internal consistency (Robot = .75; Human = .57[3]).

- **The Emotional Bond to the Interviewer**. These items assessed the degree to which the participant felt a bond was formed with the interviewer. Items were "I liked the interviewer", "It was easy to engage with the interviewer", "I felt indifferent to the interviewer" (reversed), "I communicated easily with the interviewer", "It was an awkward interaction" (reversed), "I was interested in the interviewer's words and actions", "I understood what the interviewer was asking", "I felt in rapport with the interviewer", "I felt comfortable with the interviewer" and "The interviewer was annoying" (reversed). The items demonstrated high internal consistency (Robot = .88; Human = .86).

- **Perceived Rapport Actions of the Interviewer**. This measure included items assessing the degree to which the participant perceived the interviewer to have "Acted in a friendly manner", "Paid attention to me", "Acted indifferent to me" (reversed), "Communicated easily with me", "Used actions and words that were awkward with my actions and words", "Showed interest in my words and actions", "Understood what I was trying to say", "Showed rapport with me", "Was likable", "Was controlling of me" (reversed), "Was considerate", and "Was aggressive" (reversed). The items demonstrated high internal consistency (Robot = .77; Human = .73).

[3] The Optimal Experience measure shown here has been used in previous studies relating to PD, so although its Cronbach's alpha score is lower here than desirable, this measure has been validated well enough to warrant use.

Six additional five-point Likert scaled items were included in the questionnaires assessing the robot interviewer: "Was the robot's behavior natural", "Was the robot's behavior human-like", "Did the robot look human-like", "Was the robot's speech natural", "Was the robot's size appropriate", and "Was the robot creepy". These items were examined individually without aggregation.

4.5 Data Analysis

We first analyzed descriptive statistics (means and standard deviations) on the robot questionnaires to assess the distribution of responses to the robot on the five-point Likert scale, to determine if participants were evaluating the robot on the more positive or negative evaluative end of the scale. Second, we conducted preliminary analyses to determine if there were any effects of the order or sequence in which the robot conducted its interview with the participant. Mixed repeated measures and between ANOVAs demonstrated that there were no effects of the repeated measure of order (first versus second interview) and the between factor of sequence (robot interviewer first versus the robot interviewer second) on the participants' responses to the robot and human interviewers on the measures of positive mood, optimal experience, the emotional bond to the interviewer, and perceived rapport actions of the interviewer. Therefore, we conducted a simple repeated measures (paired) t-test to determine the statistical significant and effect sizes of the human versus robot differences on these measures. We also calculated the effect size d [4] for each measure to determine the standardized difference between responses to the robot relative to the human. The lower the absolute value of the effect size d, the lower the differences in the responses to robot and human interviewer. Spearman's ρ correlational analysis were used to test the associations between the demographic data, the aggregated survey items and the robot specific questionnaire items. Finally, we took a case study approach to examining individual participant eye gaze and utterance responses to the robot and human.

5. RESULTS

5.1 Reactions to the Robot

Table 2 shows the mean and standard deviation of the four aggregate measures of the participants' written questionnaires about the robot as well as the single item assessments of the robot averaged over the participants. The results from the four aggregate measures indicate that participants had a positive mood and a good overall experience, that they formed an overall positive emotional bond with the robot and that they felt the robot was attempting to create rapport with them. The means from the six robot specific questions show an overall positive reaction as well. 69.2% of participants gave the robot the lowest possible score of 1 on the creepy measure. Participants also rated the robot's behavior and speech as fairly natural and rated the robot's size as appropriate. The only ratings that the robot received below the middle/mean value of 3 on the Likert scale were on the measures of "Was the robot's behavior human-like?" and "Did the robot look human-like?".

	Robot		**Human**	
Measure	**Mean**	**SD**	**Mean**	**SD**
Mood	4.22	.72	4.61	0.45
Optimal Experience	4.17	.61	4.30	0.48
Emotional Bond	3.87	.84	4.78	0.31
Rapport	3.79	.65	4.45	0.36
Behavior Natural	3.23	1.24	n/a	n/a
Behavior Human-Like	2.92	1.19	n/a	n/a
Look Human-Like	2.38	1.26	n/a	n/a
Speech Natural	3.25	1.22	n/a	n/a
Size Appropriate	3.15	1.21	n/a	n/a
Robot Creepy	1.62	1.19	n/a	n/a

Table 2: This table shows the ratings of the robot and human interviewers on a 5-point Likert scale averaged over all participants.

5.2 Comparisons to Human Interviewer

To test the differences between participant experiences in the robot and human interview, paired sample t-tests were calculated on each of the four aggregated variables described in the Measures section: Mood, Optimal Experience, Emotional Bond to the Interviewer and Perceived Rapport Actions. Findings are reported if the effect size d for the t-test was greater than .50. These tests found that the participants were in a less positive mood with the robot relative to the human ($t(12) = -3.03$, $p = .10$, $d = -.84$). There was not a significant difference between the participant's Optimal Experience ratings for the robot and the human ($t(12) = -1.00$, $p = .34$, $d = -.28$). A significant difference was found for the ratings of the Emotional Bond to the Interviewer ($t(12) = -3.77$, $p = .003$, $d = -1.05$) and for the ratings of Perceived Rapport Actions by the Interviewer ($t(12) = -3.48$, $p = .005$, $d = -.96$).

We conducted paired sample t-test for non-aggregated individual items on all measures in order to elaborate the meaning of the findings on composite measures. The participants, on average, rated themselves as more indifferent to the robot than to the human ($d = .82$, $p = .12$) and reported that the robot showed less interest in the participants' words and actions ($d = -.80$, $p = .01$). However, for some of the questionnaire items, there was essentially no difference between the robot and the human. If a variable had an effect size under 0.1, there was one tenth of a standard deviation difference between the robot and human, and we considered it to be a small difference. Three non-aggregated questionnaire items were under this threshold. The participants felt similarly low levels of tenseness ($d = .08$), embarrassment ($d = .000$), and difficulty concentrating ($d = .07$) during the robot and human interviews.

The correlations between the aggregate measures and the robot-only findings are reported in Table 3. Believing that the robot acted naturally was positively correlated with feeling an emotional bond with the robot ($p = .005$) and with perceiving the robot to have performed rapport actions ($p = .01$). Perceiving the robot as creepy was negatively correlated with creating an emotional bond with the robot ($p = .003$).

	Mood	**Optimal Experience**	**Emotional Bond**	**Perceived Rapport Actions**
Behavior Natural	0.21	−0.09	0.72**	0.66*
Behavior Human-Like	−0.20	−0.41	0.49	0.32
Look Human-Like	−0.11	0.10	−0.05	0.16
Speech Natural	0.26	0.35	0.18	0.49
Size Appro.	−0.14	−0.42	0.34	0.50
Robot Creepy	−0.27	0.03	−0.76**	−0.41

Table 3: This table shows the correlations between the aggregate measures and the robot specific items on the questionnaire. Items marked with a star (*) were significant at the .05 level. Items marked with two stars (**) were significant at the .01 level.

5.3 Additional Preliminary Video Analysis

Videotaped interviews were examined in a preliminary manner to assess the number of times the participant gazed at the eyes or head of the interviewer and the duration of each event. Additionally we examined how much participants spoke during the interviews in terms of number of utterances and words per utterance.

The rates at which different participants gazed at the interviewers' faces and spoke to the interviewer were very different. Comparisons of the within-subjects differences showed that some participants gazed more at the robot and some gazed less at the robot. Some talked more to the robot and some more to the human. As an example, we closely examined the results for participant #2 who spoke six utterances (3.50 average words per utterance) to the robot and spoke 81 utterances to the human interviewer (6.58 words per utterance) during the Likert-scale section of the PDQ-39. During the same time period this participant gazed at the robot interviewer's face 56 times for a total time of gaze at 4 minutes 8 seconds. During the interview with the human, the participant gazed at the interviewer's face 78 times for a total time of gaze at 7 minutes 1 second.

As a comparison, we examined participant #6 who saw the questionnaires and the interviewers in the same order as participant #2. Participant #6 spoke 16 times to the robot interviewer and 4 times to the human interviewer. This participant also gazed more at the robot, spending a total of 6 minutes and 16 seconds looking at the robot's face and only 2 minutes and 59 seconds at the humans face, despite having 32 eye gaze events for both interviewers.

These two example participants represent the larger pattern found in the data. Participants #2 and #6 varied substantially in their behavior as was typical across participants. Yet, these two participants had similar opinions of the robot compared to the human and their views were similar to the other participants. Both participants had a weaker emotional bond to the robot than to the human (#2: Robot=3.80, Human=4.80; #6:Robot=3.78, Human=4.20) and felt that the robot was doing fewer rapport action than the human was (#2: Robot=2.33, Human=4.58; #6: Robot=3.08, Human=4.08). They both felt that the robot and the human were equal in providing a good overall experience (#2=5.00, #6=4.33). Participant #2 (the participant who made less eye contact with the robot and spoke less to the robot) was in an equally good mood when talking to both interviewers (5.00), while the self-rated mood of participant #6, who spoke more and made more eye contact with the robot, was only slightly lower when talking to the robot (Robot=4.20, Human=4.60). In conclusion the behavioral analysis suggested that one cannot infer from the talkativeness and eye gaze patterns of people with PD how much they like the robot, how well they connect with the robot, or how human-like they find the robot.

6. DISCUSSION

When examining the participants' reactions to the robot without comparison to the human, the reactions are positive. Participants rated the robot higher than the middle point/mean of the Likert scale on almost all measures. They were in a good mood, felt an emotional bond, felt rapport and did not feel stigmatized during the interaction with the robot. These positive reactions show that people with PD are open to working with a robot and are willing to discuss their health with a robot.

When we examined reactions to the robot interviewer as compared to the reactions to the human, however, we found that the two were viewed as dissimilar. Participants were in a better mood with the human, they felt a stronger emotional bond to the human, and they felt that the human interviewer tried to create rapport more than the robot did. As expected, we also found quite pronounced individual differences among the subjects with respect to their eye gaze and verbal behavior towards the robot compared to the human interviewer, and the within-subjects design specifically allowed us to control for these types variations. While one could take the comparison to suggest that the robot is not appropriate for the task, it is important to note that participants felt equally capable and equally non-stigmatized during both interviews and that there was no difference in how embarrassed they felt during the two interviews. These results suggest that it may be possible for the person with PD to form a working alliance with a robot, regardless whether or not a human interviewer would be preferred over a robot. A working alliance requires, at a minimum, a shared goal and a minimal affective bond (though a stronger affective bond does improve the alliance) [15]. The robot has the traits potentially needed to be an acceptable helper, just not a better helper than a human therapist.

Our results are thus overall very encouraging. For a robot whose main job is to reduce stigmatization, the fact that a population so invested in this type of interview is comfortable enough to answer serious questions about their health represents a significant success for the robot. As we stated before, it is important to note that the robot was not disliked by the participants, as it was rated above the middle value on the Likert scale on almost all measures, but it simply was not able to rise to human level – which was not expected in the first place. While it seems that participants may not have felt emotionally close with the robot interviewer, they were still willing to engage a working alliance with the robot if only for the brief time of the interview.

Thus, the present study achieved our main aim to test whether people with PD would accept a robot in a health care function. The results suggest that the robot was acceptable but was not viewed as favorably as the human. Many factors could contribute to the differences in the ratings between the robot and the human, for example physical appearance, movements, differences in positioning of the interviewer, and not the least that one is a machine and one is human. While it is not clear which of these aspects contributed to differences between ratings of the robot, these differences are not relevant for the main aim of the study to investigate the acceptability of the robot for the intended task. It is possible that repeated interactions with the robot could lead to differences in ratings once subjects become sufficiently acquainted with the robot because none of the subjects reported having interacted with robots before. Whether these ratings will be increased and the robot viewed human-like (at least with respect to the particular task) or whether they will drop relative to the current study is an empirical question that would require a longitudinal study with repeated follow-up interaction experiments.

There are very additional exciting directions for future work using robots to assist people with PD given that we have first evidence that people with PD would accept a robot mediator. For example, we can now start to develop a comprehensive robot system that is able to reduce stigmatization felt by people with facial masking by mediating the caregiver-client interactions. And when there is no caregiver available, the robot could autonomously interview people about their health status more frequently than is otherwise practically possible for health care providers. Such frequent examinations of the state of a person's PD could lead to better therapies for that person and overall better social and physical outcomes.

7. CONCLUSIONS

In this paper we presented the first HRI study investigating whether people with Parkinson's disease would be open to interaction with a robot who can administer a health status survey. The results from our study show that overall participants felt positive towards the robot and that the robot was as good as a human at maintaining dignity and avoiding embarrassment of the person with PD. This is an important result given the stigmatization people with PD often experience. While the results also showed that the robot fell short of being rated as positively as the human interviewer, the most important implication for subsequent investigations is that robots could form "working alliances" with persons with PD in the future. It is our hope that this study will contribute to the development of future autonomous robots that could serve as "mediators" in a human caregiver-client relationship and thus help to both maintain the dignity of people with Parkinson's disease while reducing the risk of their social stigmatization.

8. ACKNOWLEDGMENTS

This project was in part supported by NSF grant #IIS-1316809.

9. REFERENCES

[1] P. Briggs, M. Scheutz, and L. Tickle-Degnen. Reactions of people with parkinson's disease to a robot interviewer. In *Proceedings of the workshop on assistive robotics for individuals with disabilities at IROS*, 2014.

[2] E. Broadbent, R. Tamagawa, A. Patience, B. Knock, N. Kerse, K. Day, and B. A. MacDonald. Attitudes towards health-care robots in a retirement village. *Australasian journal on ageing*, 31(2):115–120, 2012.

[3] A. S. Buchman, P. A. Boyle, R. S. Wilson, D. A. Fleischman, S. Leurgans, and D. A. Bennett. Association between late-life social activity and motor decline in older adults. *Archives of Internal Medicine*, 169(12):1139–1146, 2009.

[4] J. Cohen. *Statistical power analysis for the behavioral sciences*. Routledge Academic, 2013.

[5] M. Csikszentmihalyi. *Flow: the psychology of optimal experience*. Harper & Row, 1990.

[6] S. Dubowsky, F. Genot, S. Godding, H. Kozono, A. Skwersky, H. Yu, and L. S. Yu. Pamm-a robotic aid to the elderly for mobility assistance and monitoring: a "helping-hand" for the elderly. In *Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on*, volume 1, pages 570–576. IEEE, 2000.

[7] J. Fasola and M. J. Matarić. A socially assistive robot exercise coach for the elderly. *Journal of Human-Robot Interaction*, 2:3–32, 2013.

[8] D. Feil-Seifer and M. Matarić. Toward socially assistive robotics for augmenting interventions for children with autism spectrum disorders. In O. Khatib, V. Kumar, and G. Pappas, editors, *Experimental Robotics*, volume 54 of *Springer Tracts in Advanced Robotics*, pages 201–210. Springer Berlin Heidelberg, 2009.

[9] C. Jenkinson, R. Fitzpatrick, V. Peto, R. Greenhall, and N. Hyman. The parkinson's disease questionnaire (pdq-39): development and validation of a parkinson's disease summary index score. *Age and Aging*, 26:353 – 357, 1997.

[10] H.-T. Jung, J. Baird, Y.-K. Choe, and R. A. Grupen. Upper extremity physical therapy for stroke patients using a general purpose robot. In *RO-MAN, 2011 IEEE*, pages 270–275. IEEE, 2011.

[11] E. S. Kim, L. D. Berkovits, E. P. Bernier, D. Leyzberg, F. Shic, R. Paul, and B. Scassellati. Social robots as embedded reinforcers of social behavior in children with autism. *Journal of Autism and Developmental Disorders*, 43:1038 – 1049, 2013.

[12] R. McNaney, J. Vines, D. Roggen, M. Balaam, P. Zhang, I. Poliakov, and P. Olivier. Exploring the acceptability of google glass as an everyday assistive device for people with parkinson's. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 2551–2554. ACM, 2014.

[13] N. M. Puccinelli and L. Tickle-Degnen. Knowing too much about others: moderators of the relationship between evesdropping and rapport in social interaction. *Journal of Nonverbal Behavior*, 28(4):233 – 241, 2004.

[14] L. Tickle-Degnen. Nonverbal behavior and its functions in the ecosystem of rapport. *The SAGE handbook of nonverbal communication*, pages 381–399, 2006.

[15] L. Tickle-Degnen. *Occupational therapy for physical dysfunction*, chapter Therapeutic Rapport, pages 412 – 427. Lippincott William & Wilkins, 7 edition, 2014.

[16] L. Tickle-Degnen and K. D. Lyons. Practitioners' impressions of patients with parkinson's disease: the social ecology of the expressive mask. *Social Science & Medicine*, 58:603 – 614, 2004.

[17] L. Tickle-Degnen, M. Saint-Hilaire, C. A. Thomas, B. Habermann, L. S. S. Martinez, N. Terrin, F. Noubary, and E. N. Naumova. Emergence and evolution of social self-management of parkinson's disease: study protocol for a 3-year prospective cohort study. *BMC Neurology*, 14, 2014.

[18] L. Tickle-Degnen, C. Thomas, M. Saint-Hilaire, E. Naumova, N. Ambady, T. Ellis, and R. Wagenaar. The social self-management of parkinson's disease (pd) in daily life. *Movement Disorders*, 27:S313–S314, 2012.

[19] L. Tickle-Degnen, L. A. Zebrowitz, and H. ing Ma. Culture,gender and health care stigma: Practitioners' response to facial masking experienced by people with parkinson's disease. *Social Science & Medicine*, 73:95 – 102, 2011.

[20] K. Werner, J. Oberzaucher, and W. Franz. Evaluation of human robot interaction factors of a socially assistive robot together with older people. In *Sixth International Conference on Complex, Intelligent, and Software Intensive Systems*, pages 455–460. IEEE, 2012.

[21] W. H. O. (WHO). International classification of functioning, disability, and health (icf). 2001.

[22] A. W. Willis, B. A. Evanoff, M. Lian, S. R. Criswell, and B. A. Racette. Geographic and ethnic variation in parkinson disease: A population-based study of us medicare beneficiaries. *Neuroepidemiology*, 34:143 – 151, 2010.