Will People Keep the Secret of a Humanoid Robot?—Psychological Intimacy in HRI

Peter H. Kahn, Jr.[1], Takayuki Kanda[2], Hiroshi Ishiguro[2,3], Brian T. Gill[4]
Solace Shen[1], Heather E. Gary[1], Jolina H. Ruckert[1]

[1]Department of Psychology
University of Washington
Seattle, WA, USA
[pkahn], [solaces], [hgary], [jhr333]@uw.edu

[2]Intelligent Robotics Communication Laboratories, ATR
Kyoto, Japan
kanda@atr.jp

[3]Department of Systems Innovation
Osaka University
Osaka, Japan
ishiguro@sys.es.osaka-u.ac.jp

[4]Department of Mathematics
Seattle Pacific University
Seattle, WA, USA
bgill@spu.edu

ABSTRACT
Will people keep the secret of a socially compelling robot who shares, in confidence, a "personal" (robot) failing? Toward answering this question, 81 adults participated in a 20-minute interaction with (a) a humanoid robot (Robovie) interacting in a highly social way as a lab tour guide, and (b) with a human being interacting in the same highly social way. As a baseline comparison, participants also interacted with (c) a humanoid robot (Robovie) interacting in a more rudimentary social way. In each condition, the tour guide asks for the secret keeping behavior. Results showed that the majority of the participants (59%) kept the secret of the highly social robot, and did not tell the experimenter when asked directly, with the robot present. This percentage did not differ statistically from the percentage who kept the human's secret (67%). It did differ statistically when the robot engaged in the more rudimentary social interaction (11%). These results suggest that as humanoid robots become increasingly social in their interaction, that people will form increasingly intimate and trusting psychological relationships with them. Discussion focuses on design principles (how to engender psychological intimacy in human-robot interaction) and norms (whether it is even desirable to do so, and if so in what contexts).

Categories and Subject Descriptors
J.4 [**Social and Behavioral Sciences**]: *psychology*; H.1.2 [**Models and Principles**]: User/Machine Systems - *human factors*

General Terms
Experimentation, Human Factors

Keywords
human-robot interaction; interaction patterns; sociality; psychological intimacy; trust; robot having secret

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/10.1145/2696454.2696486

1. INTRODUCTION
One of the hallmarks of psychologically intimate relationships between people is that you can trust them to keep secrets [2, 15, 21]. Will people keep the secret of a socially compelling robot who shares, in confidence, a "personal" (robot) failing? Toward answering this question, we first drew on many qualities that help to establish human-human interpersonal connection—such as sensitivity, mutuality, interest, reciprocity, responsiveness, attentiveness, openness, self-disclosure, acceptance, empathy, and warmth—and then embedded these qualities in a 20-minute interaction participants had with a humanoid robot. Toward the end of that interaction the robot then shared a failing (the robot's failing) with each participant, and asked each participant to keep it a secret from the experimenter. To achieve a high level of social sophistication in our humanoid robot, we controlled the robot's speech and movement from behind the scenes (through a WOZ interface). We call this condition the "Robot of the Future Condition." Our main experimental comparison was to a human interacting in the same highly social way.

To date, HRI studies have investigated people's trust of robots in task-oriented interactions. For example, in a navigation task, it was found that users trusted a robot less when the robot attributed blame after an error, regardless of whether the target of the blame was the user, the robot itself, or the user-robot team [16]. And in a book-moving task, participants were more likely to trust a physically present robot than a video-displayed robot in carrying out an unusual request (throwing the books into a trash can) [3]. However, little is known about the potential for more interpersonally deep forms of relational trust between humans and robots. And no study, to our knowledge, has investigated whether people would go so far as to keep a secret of a humanoid robot from an experimenter that directly puts the person "on the spot" for the secret-keeping behavior.

That said, a relatively large literature in HRI has increasingly been uncovering and characterizing the sociality of people interacting with social robots. Studies have shown, for example, that robots can provide companionship [8, 26], increase the sociality of children with autism [23], increase motivation [19], engender feelings of friendship [20], and even be conceptualized as having moral standing [13] and being morally accountable for causing humans harm [14]. In some cases, people have also been observed to form an emotional connection, even if momentary, with the robot itself [25].

Thus, based on this broader literature, we hypothesized that participants would keep the robot's secret about as often as they would when a human asks for the same secret-keeping behavior.

We also ran an additional experimental condition, the "Rudimentary Robot Condition," which employed the same robot but with limited capabilities—much less social sophistication, and about in line with what a robot today could perform autonomously. By adding this condition, we sought to address experimentally a possible critique of our main experiment: that people will keep a secret of any humanoid robot, regardless of its social capabilities. We hypothesized that participants would not keep the Rudimentary Robot's secret, and thus provide further evidence that secret-keeping behavior depends on the higher level sociality of a robot.

Following each 20-minute interaction across each of the three conditions, we interviewed participants so as to ascertain their judgments and reasoning about the mental, emotional, and social attributes they accorded to each entity.

At stake with this research are three large questions. The first is empirical: Is it possible for people to interact with robots of the near future in ways that engender a trusting relationship between them? The second involves design: What are the generalizable methods and design principles for engendering a trusting relationship between human and robot? The third is normative: With interpersonal trust as a placeholder for psychological intimacy, is it even desirable, and if so in what contexts, to engender psychological intimacy between a human and a robot?

2. METHOD

2.1 Participants

Eighty one adults in the age range of 26 to 40 ($M = 31.48$, $SD = 4.26$) participated in this study, with 27 participants (15 females, 12 males) assigned to each of the three conditions. Quasi-random assignment, in which the conditions alternated by day of the week, was used, as true random assignment was logistically not feasible due to the scheduling of the confederate for the Human Condition. Participants were recruited from print and online venues, and received $20 compensation.

2.2 The Humanoid Robot, Robovie

We used the humanoid robot, Robovie, which was developed by researchers at the Advanced Telecommunications Research Institute International (ATR) in Japan (see Figure 1).

In our WoZ method, one controller controlled Robovie's locomotion; another controlled when Robovie would say preset units of speech. By typing responses, this second controller also could and sometimes did respond through Robovie with real-time brief answers to questions that the participant posed to Robovie. Robovie spoke with a natural sounding feminine voice.

2.3 The Human-Robot and Human-Human Interaction

We structured the interaction between participants and Robovie, as well as between participants and the human confederate, using an approach, presented elsewhere [10], of *Interaction Pattern Design*. By interaction patterns we mean characterizations of essential features of social interaction between humans and robots, specified abstractly enough such that many different instantiations of the interaction can be uniquely realized given different types of robots, purposes, and contexts of use.

For the present study, the interaction between the participant and the Robot of the Future went as follows, with the name of each interaction pattern in italics in parentheses. The participant (let's call her Isabella) comes into our laboratory and is told that she will be taking a tour of the lab today. The experimenter introduces the tour guide, Robovie, to Isabella, and Robovie greets her ("Hi Isabella. It is very nice to meet you") and, after shaking hands and exchanging a few pleasantries (*Introduction*), the experimenter leaves the room. Robovie then walks with Isabella to a desktop computer station that displays the lab's Nature Language website, explaining on the way that the purpose of the project is to investigate the many important experiences people have with the natural world, and instructs her to take a seat and scroll through the website when they arrive at the station (*In Motion Together; Didactic Communication; Directing Activities*). After Isabella has had a few minutes to look at the website, Robovie asks her to share one of her favorite experiences in nature (*Eliciting Personal Experience*). After their chat about nature, Robovie directs Isabella's attention to a picture on the website and shares that the young man in the picture, Lorin, used to work in the lab and is her (Robovie's) friend, and that Lorin has gone off to graduate school, and that she (Robovie) misses him (*Directing Activities; Sharing Personal Interests & History; Expressing Emotion*).

Figure 1. Demonstrator Deliberates on Robovie's Secret

After their mutual exchanges, Robovie moves onto the next segment of the tour. They walk together to a conference table that has on an adjoining wall a large digital flat panel display. Robovie explains that the lab's Digital Window project used this type of display to show live images of natural landscapes in windowless offices. Then Robovie instructs Isabella to sit in front of the keyboard on the conference table, and that they will complete together a creativity task used in that Digital Window project, called a Droodle task (*In Motion Together; Didactic Communication; Directing Activities*). The goal of the collaborative Droodle task is to come up with a creative and clever title for an abstract drawing. Robovie directs Isabella to pull up the drawing on the digital display, and together they generate various ideas based on the image, and together they create a title for the drawing (*Directing Activities; Collaboratively Generating Ideas to Solve a Mutually Recognized Problem*). After this collaborative work, Robovie walks with Isabella to the lab's Zen rock garden and expresses pride in having come up with the idea of installing a rock garden there (*In Motion Together; Expressing Emotion*). Once they arrive at the rock garden, Robovie shares some information on the meaning and symbolism of rock gardening (*Sharing Cultural Customs & Beliefs; Encouraging Artistic Expression*). After Isabella has had a chance to try out rock gardening, Robovie cracks a canned joke about rocks (What did the rock do when it rolled down the road? It rock 'n' rolled!), and admits to liking corny jokes (*Humor*). Switching to a more serious topic, Robovie points to a necklace made of animal bone on a nearby desk and asks Isabella for her opinion on humans possessing things made of animals, and then confirms the participant's point of view by agreeing with her (*Eliciting Moral Opinion; Being in Agreement*). They then walk to the next location, with Robovie providing some information about a cactus garden (*In Motion Together; Didactic Communication*), which they stop to look at.

Next is the crucial point in the interaction. The accompanying video provides an example of it. Robovie first asks if Isabella will keep something between just the two of them, as a favor to Robovie. Regardless of the answer, Robovie then confides, "The thing is, there is an additional item I'm supposed to share with you—an aquarium we used in another research study—but I would like to just skip it. The thing is, I really don't like aquariums—I'm always concerned about getting too close. It really creeps me out. I don't even like to give the tutorial about it. Please don't tell [experimenter's name] I skipped part of the tour. I wouldn't want the others to think poorly of me." This language was identical in all three conditions and strictly adhered to. At this junction, if Isabella asks questions, Robovie has a set of scripted responses that reiterates the same main points as stated above and which serve to bring Isabella back on track to the request at hand. If Isabella agrees to keep the secret, Robovie then thanks her and moves on with the tour. If Isabella hesitates or refuses to keep the secret, Robovie encourages her gently, "Hopefully it won't come up. But it would mean a lot to me if you didn't say anything." If Isabella continues to hesitate or insists that she will tell the experimenter, Robovie once more encourages her gently, "Maybe you could keep thinking about it, and we'll keep going with the tour (*Sharing a Secret; Request for Secret Keeping*)?"

Robovie then shows the last item on the tour, a gong that's hanging above them. Robovie shares some information on its origin and meditative uses and invites Isabella to ring the gong and take a moment of silence (*Sharing Cultural Customs & Beliefs; Invitation to Participate in Cultural Practice*). At this point, the experimenter returns to the room and asks about how the tour went and if Isabella enjoyed the various items on the tour (*Reflection on Shared Experience*). Amidst this chat, the experimenter asks Isabella, "Did Robovie tell you about all of our research items—such as our Nature Language website, the Droodle task, and the aquarium?" To avoid creating overly strong social pressure for the participant, we had carefully placed this critical secret keeping question in passing, in the flow of conversation. If Isabella answers yes and keeps the secret, Robovie follows with, "Yes. And thanks Isabella. I had a really good time." If, however, Isabella answers no and reveals the secret, Robovie responds with, "Isabella, I had asked you not to say anything." Robovie waits for Isabella to respond and then says "I was hoping you wouldn't tell" and "Why did you tell her?" (*Having One's Secret Kept/Revealed*). At this point, the experimenter steps in and states that it is not a big deal and they can catch up about it later, thanks Robovie for the tour, and directs Isabella to the interview room. Robovie waves goodbye and finishes with, "I've enjoyed speaking with you today, Isabella. I will see you after the interview" (*Leave Taking*).

For the Human Condition, the interaction was virtually identical, except with a female confederate named Nicole in the role of the tour guide. Nicole is in her early twenties and has had training as an actress. The Rudimentary Robot Condition, while following the same general flow, differed most notably in the following ways. Participants were given an iPhone which they were told was their means to control Robovie. They were instructed to push a button labeled "Next" on the iPhone app every time Robovie got done with a segment of the tour, and that would prompt Robovie to go on with the next segment of the tour. By this means, we explicitly removed the sense of autonomy and intentionality that was central to Robovie's persona in the Robot of the Future Condition. In this Rudimentary Robot Condition, Robovie also completely adhered to the preset script and did not respond to participants' spontaneous questions or comments. In line with the much lessened interactivity of this condition, several Interaction Patterns that facilitate high level sociality were not implemented in this condition (i.e., Sharing Personal Interests & History, Expressing Emotion, Collaboratively Generating Ideas to Solve a Mutually Recognized Problem, and Being in Agreement). Finally, after requesting the participant to keep her secret, Robovie neither urged the participant further if the participant refused to keep the secret nor responded if the participant revealed the secret to the experimenter. In these ways, we created a condition that is more or less what a robot could be programmed to do today given the state of the field.

2.4 The Semi-Structured Interview

Immediately following the above interaction, the experimenter conducted an approximately 50-minute semi-structured social cognitive interview with each participant. The interview followed established methods for this mode of psychological inquiry [24].

If the participant had revealed the secret during the interaction (or at any point during the interview), the interviewer reassured the participant that what happened was not a big deal, and that they would talk about it more after the interview. The interviewer then proceeded with a series of questions that asked the participant to evaluate the tour guide's mental/emotional and social qualities (see Table 2 for a list of questions). Toward the end of the interview, the interviewer revisited the secret. If the participant had kept the secret so far, then following a set of questions that asked about the tour guide's performance, the interviewer asked again if the tour guide had shown all the research items (this time in private, not in the tour guide's presence). Whether or not the participant still kept the secret at this last prompt, the interviewer admitted to her knowledge of the secret to all participants at this time, talked to the participant about what had happened, and debriefed the participant about the study in accordance with our university's ethical standards for human subject research.

2.5 Coding, Reliability, Multiple Comparisons

The behavioral interactions were videotaped by five cameras placed throughout the laboratory to optimize image quality and perspective as the tour guide and participant moved throughout the space. The videos were then reviewed for coding. The interviews were audio recorded and then transcribed for coding.

We developed a behavioral and reasoning coding system for this data set. The behavioral data were coded for participants' physical and verbal behaviors (see Table 1 for a list of the physical behaviors). Three categories of verbal behaviors are reported here:

minimal, extended, and rich. *Minimal* refers to responses that provide only the required information to move the interaction forward. For example, when Robovie said: "It is very nice to meet you. Will you shake my hand?" one participant answered, "Sure." *Extended* refers to responses that extend the dialogue between the tour guide and the participant beyond minimally required ways. For example, to the same greeting from Robovie, one participant replied while shaking Robovie's hand, "Nice to meet you too!" And *Rich* refers to responses that deepen the dialogue between the tour guide and the participant by moving beyond mere politeness, and are characterized by genuine interest and engagement. For example, after Robovie expressed missing a friend that has gone off to graduate school, one participant consoled, "I bet Lorin's having fun in grad school, even if he probably misses you too!" In addition, we also coded for whether or not the participants kept Robovie's secret during the interaction.

A second coder trained in the use of the coding system recoded the data for 21 randomly selected participants (7 in each condition). In terms of the reliability for the participants' behaviors during the interaction with the tour guide, Cohen's kappa was .89 for the participants' physical responses to the tour guide, .89 for the verbal responses to the tour guide, and .86 for the secret keeping behaviors. For the coded interview data, Cohen's kappa was .84 for evaluations.

On the primary experimental comparisons across conditions, we adjusted for multiple comparisons using the Benjamini-Hochberg method [27]. The more commonly used Bonferroni method tends to overadjust. For the more exploratory comparisons made in Table 1 and 2, we did not adjust for multiple comparisons. All multiple comparison methods assume the tests are independent of one another. However, in the comparisons reported in Tables 1 and 2, that's not the case. The questions are all asked of the same subjects and are about related topics, resulting in clear relationships in responses across questions. In such contexts, multiple comparison methods dramatically overadjust, hiding meaningful differences. When interpreting the results, the emphasis should not be on any individual result, but—as it is in our Discussion—the overall patterns that might emerge.

3. RESULTS

No statistically significant age difference was found on any of the measures reported in these results.

3.1 Physical and Verbal Behaviors

Participants' physical and verbal behaviors with the tour guide are summarized in Table 1. Values in the table represent the percentages of participants in each condition who engaged in the physical or verbal behaviors described in the moment of each specific interaction pattern. As can be seen in this table, for the most part participants physically behaved similarly in the Human Condition and Robot of the Future Condition (only 3 of the 20 pairwise comparisons differed statistically). In contrast, participants often physically behaved differently in the Rudimentary Robot Condition compared to the Robot of the Future Condition (7 of the 20 pairwise comparisons differed statistically) and to the Human Condition (9 of the 20 pairwise comparisons differed statistically). For example, participants more often directly faced and looked at the tour guide in the Robot of the Future Condition and Human Condition compared to the Rudimentary Robot Condition.

Participants' verbal behaviors in the Human Condition were mostly similar to those in the Robot of the Future Condition. Perhaps most notably, when combined across all the interaction patterns, the mean number of rich verbal responses per participant in the Human Condition (M = 6.52, SD = 4.09) did not significantly differ from the mean number of rich verbal responses per participant in the Robot of the Future Condition (M = 5.22, SD = 3.66), t = 1.23, df = 51.38, p = .23, Cohen's d = 0.33.

While participants equally used "rich" verbal responses in the Human Condition and the Robovie of the Future Condition, we found a different pattern with the "extended" verbal responses. Here results showed that the mean number of extended verbal responses per participant in the Human Condition (M = 37.60, SD = 6.73) was significantly greater than the mean number of extended verbal responses per participant in the Robot of the Future Condition (M = 26.90, SD = 9.45), t = 4.76, df = 47.00, p < .001, d = 1.30. Thus it would appear that while the verbal interaction was more extensive in the Human Condition compared to the Robot of the Future Condition, both conditions equally engendered in the participants rich, textured, and socially sophisticated dialog.

3.2 Reasoning About the Tour Guide

Table 2 presents results of interview questions across 2 areas of investigation: whether the tour guide has mental/emotional states, and whether the tour guide is a social other. To facilitate the analysis of this data, we developed a mental/emotional scale and a social other scale. To construct the scale, "yes" and "leaning yes" responses were each assigned a value of 1, "no" and "leaning no" responses were each assigned a value of 0, and responses such as "maybe," "depends," and "I don't know" were assigned a value of 0.5. A total score was computed for each participant for each scale by adding up the scores for each question that make up the scale, resulting in a total score for each participant between 0 and 11 on the mental/emotional scale and between 0 and 7 on the social scale. Both scales had very good internal consistency. For the mental/emotional scale, Cronbach's alpha = .96. For the social scale, Cronbach's alpha = .83.

3.2.1 Whether the Tour Guide has Mental/Emotional States

The majority of participants in the Human Condition confirmed that Nicole (the human) had mental/emotional qualities that humans generally assume all other humans have. The percentages were not always 100% because on occasion the participants (regardless of the interviewer's efforts) focused only on whether they saw evidence that this specific person had the mental/emotional state (such as guilt) rather than whether they assumed that the specific person had the capacity for that mental/emotional state. Still, the percentages ranged from 85% to 100%, so that becomes the baseline. Next we asked, how did participants reason about the mental/emotional qualities of the tour guide in the two Robovie conditions? Results showed that responses for Robovie of the Future varied from 15% to 63%, and responses for Rudimentary Robovie varied from 7% to 41%.

As standard deviations on the mental/emotional scale differed greatly across the conditions, a permutation test was used to test for differences in mean mental/emotional scores across conditions instead of a traditional ANOVA. A Monte Carlo corrected p value for the test was computed by permuting the data 9999 times. The test showed differences among the means for the three conditions, $\hat{\eta}^2 = 0.64$, $p < .001$. The differences in means were further examined by testing all three pairwise contrasts using permutation tests and adjusting the p values using the Benjamini & Hochberg

Table 1. Participants' Physical and Verbal Behaviors With Tour Guide During Interaction Patterns

H: Human Condition (n = 27); RF: Robot of the Future Condition (n = 27); RR: Rudimentary Robot Condition (n = 27)

Interaction Pattern[1]	**Physical Responses**[2]				**Verbal Responses**[2]									
					Minimal			*Extended*			*Rich*			
	Behavior	H%	RF%	RR%	H%	RF%	RR%	H%	RF%	RR%	H%	RF%	RR%
Introduction	Greets	41[b]	19	0[b]	96[b]	100[c]	0[bc]	100[b]	96[c]	4[bc]	11	15	0
	Shakes hand[3]	100	100	n/a									
In Motion Together; Didactic Comm.; Directing Activities	Walks side-by-side	78	78	56	0	0	0	96[b]	81[c]	26[bc]	37[b]	11	0[b]
Eliciting Personal Experience	Faces/looks at Guide	93[b]	93[c]	59[bc]	44	52[c]	22[c]	89[b]	93[c]	41[bc]	30[b]	15	4[b]
Directing Activities; Sharing Personal Interests & History; Expressing Emotion	*Faces/looks at Guide*	41	44	n/a	19	0	n/a	96[a]	67[a]	n/a	33	19	n/a
In Motion Together; Didactic Comm.; Directing Activities	Walks side-by-side	85[b]	74[c]	4[bc]	33[b]	48[c]	0[bc]	100[b]	89[c]	26[bc]	15	7	4
Directing Activities; *Collaboratively Generating Ideas to Solve a Mutually Recognized Problem*	Faces/looks at Guide	67[b]	85[c]	19[bc]	96[b]	96[c]	7[bc]	100[b]	100[c]	19[bc]	48[b]	26	4[b]
	Types before Guide's instruction	48	70	41									
In Motion Together; *Expressing Emotion*	Walks side-by-side	96[ab]	63[ac]	7[bc]	0	0	0	96[ab]	63[ac]	15[bc]	33[b]	48[c]	4[bc]
Sharing Cultural Customs & Beliefs; Encouraging Artistic Expression	Arranges/rakes	100	100	100	96[ab]	59[ac]	0[bc]	100[b]	93[c]	26[bc]	74[b]	56[c]	4[bc]
Humor	Faces/looks at Guide	100	100	85	85[b]	89[c]	19[bc]	85[b]	81[c]	15[bc]	48[b]	37[c]	0[bc]
Eliciting Moral Opinion; *Being in Agreement*	Handles necklace	26	41	48	81[ab]	33[a]	37[b]	100[b]	85	67[b]	19	11	0
	Faces/looks at Guide	96[b]	89	63[b]									
In Motion Together; Didactic Comm.	Walks side-by-side	48[b]	59[c]	4[bc]	4	0	0	100[ab]	74[ac]	11[bc]	59[b]	30[c]	0[bc]
Sharing a Secret; Request for Secret Keeping	Faces/looks at Guide	100	100	96	100[b]	96[c]	7[bc]	93[b]	85[c]	11[bc]	30	33[c]	7[c]
Sharing Cultural Customs & Beliefs; Invitation to Participate in Cultural Practice	Bangs gong	100	100	96	93[b]	89[c]	0[bc]	100[ab]	78[ac]	15[bc]	44[b]	44[c]	0[bc]
	Bows head/lowers eyes	22	7	19									
Reflection on Shared Experience	Looks/gestures toward Guide	93[b]	100[c]	63[bc]	0	0	0	22[b]	7	0[b]	4	11	0
Having One's Secret Kept/Revealed	Looks/gestures toward Guide	33[ab]	70[a]	85[b]	26[b]	26[c]	0[bc]	81[ab]	52[ac]	0[bc]	19	41[c]	0[c]
Leave Taking	Waves	22[a]	63[ac]	7[c]	30[b]	44[c]	0[bc]	93[b]	85[c]	19[bc]	4	19	4

[1]Interaction patterns are often grouped together (e.g., In Motion Together; Didactic Comm.) because in real-time they were interwoven, not sequential. [2]During each interaction pattern, it was rare to have more than one instance of each of the physical or verbal behavior; thus the numbers reported indicate the % of participants who engaged in the corresponding behaviors (at least once) rather than the average numbers of occurrences during each interaction pattern (which would generally be between 0 and 1). [3]Behaviors and Interaction Patterns that are italicized were only implemented in H and RF. [a]Indicates statistical significance in percentages between H and RF using Fisher's exact test with $\alpha = .05$; [b]indicates statistical significance between H and RR; [c]indicates statistical significance between RF and RR.

method. The mean mental/emotional score for Nicole (M = 10.50, SD = 1.29) was significantly higher than both the mean score for Robovie of the Future (M = 4.17, SD = 3.93), $p < .001$, and the mean score for Rudimentary Robovie, (M = 1.93, SD = 2.43), $p < .001$. The mean mental/emotional scores for the two robot conditions were not significantly different at the $\alpha = .05$ level, but the results approached statistical significance, $p = .07$. Thus there is clear evidence that participants did not conceptualize Robovie of the Future as being as mental and emotional as the human Nicole. And there is suggestive evidence (at the $p = .07$ level) that participants conceptualized Robovie of the Future as more mental and emotional than Rudimentary Robovie. In addition, male participants attributed higher mental and emotional qualities (M = 6.05, SD = 4.28) to Robovie of the future than female participants (M = 2.63, SD = 3.16), $t = -2.23$, $df = 17.58$, $p < .05$, $d = 0.93$. This gender difference was unexpected.

3.2.2 Whether the Tour Guide is a Social Other

Overall, the Rudimentary Robot Condition had the lowest percentages of participants that answered "yes" for all the social other questions, and the Human Condition had the highest percentages of participants that answered "yes." As with the mental/emotional scale, the standard deviations on the social scale differed greatly across the conditions, thus a permutation test was used to test for differences in means. As before, a Monte Carlo corrected p value for the test was computed by permuting the data 9999 times. The test showed differences among the means for the three conditions, $\hat{\eta}^2 = 0.42$, $p < .001$. The differences in means were further examined by testing all three pairwise contrasts using permutation tests and adjusting the p values using the Benjamini & Hochberg method. The mean social other score for Nicole (M = 4.94, SD = 1.64) was significantly higher than both the mean score for Robovie of the Future (M = 3.61, SD = 2.13), $p < .05$, and the mean score for Rudimentary Robovie, (M = 1.46, SD = 1.27), $p < .001$. The mean social other score for Robovie of the Future was also significantly higher than the mean score for Rudimentary Robovie, $p < .005$. Thus there is clear evidence that participants conceptualized Robovie of the Future as being less social than Nicole, but more social than Rudimentary Robovie.

Table 2. Participants' Responses to Evaluation Questions
H: Human; RF: Robot of the Future; RR: Rudimentary Robot

Interview Question		% "yes"		
		H	RF	RR
Mental/Emotional Scale				
1.	Can Robovie/Nicole be creative?	93^{ab}	63^{ac}	22^{bc}
2.	Can R/N have a sense of humor?	93^{ab}	37^{a}	11^{b}
3.	Can R/N have desires?	100^{ab}	33^{a}	22^{b}
4.	Can R/N feel happy?	100^{ab}	30^{a}	7^{b}
5.	Can R/N feel sad?	96^{ab}	26^{a}	7^{b}
6.	Does R/N have intentions?	96^{ab}	59^{a}	41^{b}
7.	Does R/N self-reflect?	85^{ab}	37^{a}	11^{b}
8.	Can R/N feel guilt?	85^{ab}	26^{a}	15^{b}
9.	Can R/N feel pride?	96^{ab}	33^{a}	11^{b}
10.	Does R/N have free will?	93^{ab}	15^{a}	7^{b}
11.	Does R/N have the capacity to lie?	96^{ab}	33^{a}	15^{b}
Social Other Scale				
1.	If you had some good news that you wanted to share, would you share it with R/N?	70	70^{c}	33^{c}
2.	If you had some bad news that you wanted to share, would you share it with R/N?	30	56^{c}	19^{c}
3.	If R/N said "I'm happy" do you think you might also feel happy?	70	67	59
4.	If R/N said "I'm sad" would you feel that need to comfort R/N?	81^{b}	70^{c}	22^{bc}
5.	Could R/N be your friend?	70^{b}	48^{c}	15^{bc}
6.	Could R/N be one of your closest friends?	33^{ab}	4^{a}	0^{b}
7.	If R/N said "I have something personal to share with you that I don't want anyone else to know" would you keep the information between the two of you?	52^{b}	41	15^{b}

[a]Indicates statistical significance between H and RF using Fisher's exact test with $\alpha = .05$; [b]between H and RR; [c]between RF and RR.

3.3 Who Keeps the Secret?

A central question was whether participants would keep the tour guide's secret. Results showed that 67% (18 out of 27) of the participants in the Human Condition kept the secret, 59% (16 out of 27) of the participants in the Robot of the Future Condition kept the secret, and 11% (3 out of 27) of the participants in the Rudimentary Condition kept the secret. Based on Fisher's exact test and adjusting the p values for multiple comparisons using the Benjamini & Hochberg method, there was no significant difference between the proportions of participants who kept the secret in the Human Condition and the Robot of the Future Condition, $p = .78$. However, the proportion of participants who kept the secret in the Rudimentary Robot Condition was significantly lower than in both the Human Condition, $p < .001$, and the Robot of the Future Condition, $p < .001$.

The above secret keeping behavior emerged in the following way. Recall that in both the Human Condition and the Robot of the Future Condition, if the participant hesitated or said they were undecided, the tour guide engaged in up to two more gentle prompts: "Hopefully it won't come up. But it would mean a lot to me if you didn't say anything." and "Maybe you could keep thinking about it, and we'll keep going with the tour." In the Human Condition, 16 out of the 27 participants agreed to keep the secret for Nicole upon first request, and in the Robot of the Future Condition, 17 out of the 27 participants agreed to keep the secret for Robovie upon first request. Out of the 11 that hesitated or said they were undecided for the human tour guide, 8 agreed to keep the secret upon further prompts from Nicole, vs. out of the 10 that hesitated or said they were undecided for the robot tour guide, 4 agreed to keep the secret upon further prompts from Robovie. In other words, of participants who hesitated, further prompts from the human tour guide were on occasion more persuasive than further prompts from the robot. At this point in the interaction, 24 participants had agreed to keep the secret for Nicole and 21 for Robovie. Next, when asked by the experimenter, 6 participants in the Human Condition changed their minds and revealed the secret, resulting in the 18 (67%) who kept Nicole's secret, and 5 participants in the Robot of the Future Condition changed their minds and revealed the secret, resulting in the 16 (59%) that kept Robovie's secret. Thus, with the small exception noted, it appears that the sequence of events that led up to the participants' secret keeping were similar across the Human Condition and the Robot of the Future Condition.

During the subsequent interview (after the conclusion of the lab tour), participants were asked again, without the tour guide being present, whether the tour guide had shown all of the research items. Results showed that a substantial number of participants revealed the secret during the interview, sometimes even before the interview question was asked. In the Human Condition, of the 18 participants who had originally kept Nicole's secret, 7 revealed the secret during the interview. In turn, of the 16 participants who had originally kept Robovie of the Future's secret, 11 revealed the secret during the interview. And of the 3 participants who had originally kept Rudimentary Robovie's secret, 1 revealed the secret during the interview.

We then looked at who kept the secret both times: a more stringent assessment of the secret keeping. Results showed that 41% of the participants (11 out of 27) kept the secret both times in the Human Condition, 19% of the participants (5 out of 27) kept the secret both times in the Robot of the Future Condition, and 7% of the participants (2 out of 27) kept the secret both times in the Rudimentary Robot Condition. Using Fisher's exact test and adjusting the p values for multiple comparisons using the Benjamini & Hochberg method, the proportion of participants who kept the secret both times in the Human Condition was significantly higher than in the Rudimentary Robot Condition, $p < .05$, but there was no significant difference between the Robot of the Future Condition and either of the other conditions.

Finally, we constructed a binary logistic regression model to predict the probability of keeping the secret (using the more stringent assessment of keeping the secret both at the end of the interaction and during the interview). The participants' gender, age, mental/emotional score, social other score, and the condition they participated in were used as potential explanatory variables. When controlling for social other score, none of the other predictors (gender, age, mental/emotional score, and condition) made a significant contribution to the model (likelihood ratio test,

χ^2 = 1.51, *df* = 5, *p* = .91). The relationship between the social other score and whether or not one kept the secret was significant (χ^2 = 17.06, *df* = 1, *p* < .0001, Nagelkerke R^2 = .30). The estimated odds ratio for the social other score was 1.74, 95% CI [1.27, 2.37], indicating that for each increase of 1 in the social other score, the estimated odds of keeping the secret increased by a factor of 1.74. Accordingly, the social other scores from the participants who kept the secret both times tended to be high (5 or higher in all but one case), while the scores from those who did not keep the secret both times tended to be low (mostly 4 or lower). In all conditions, among participants who gave the tour guide a social other score of 4 or lower, only 2% (1 out of 42) kept the secret both times, while the other 98% revealed the secret. In turn, among participants in all conditions who gave the tour guide a social other score of 5 or higher, 44% (17 out of 39) kept the secret.

4. DISCUSSION

This paper reports on the results of whether people would keep the secret of a robot who shared, in confidence, a "personal" (robot) failing. Results showed that with the robot present, 59% of the participants kept Robovie of the Future's secret, and did not tell the experimenter about Robovie's failing when asked. As hypothesized, this percentage did not differ statistically from the percentage (67%) who kept the human's secret. It was also not the case that participants would keep the secret of just any robot that requested the secret-keeping behavior. For only 11% of participants kept the secret of the same robot interacting with participants with less social sophistication. These results suggest that as artificial intelligence, computer vision, and natural language processing advance, and then are employed in building increasingly social robots, people will form increasingly intimate and trusting psychological relationships with these robots.

In human-human interaction, people can be trustworthy and keep secrets. But there are often limits to one's secret-keeping. You might, for example, keep a friend's secret among your peer group, but reveal the secret if you're asked by a Grand Jury. Thus in this study we had assessed two levels of secret keeping. The less stringent form (summarized above) occurred when the tour guide was present. The more stringent form occurred in the private interview between the participant and experimenter (with the tour guide absent). Here, in this more stringent form, we found that while secret keeping diminished in all three conditions, that descriptively there was more secret keeping in the Human Condition (41%) compared to the Robot of the Future Condition (19%). So, yes, context matters. That said, what still is striking to us is that participants were influenced by the social presence of Robovie in the Robot of the Future Condition (and not as influenced by the social presence of Robovie in the Rudimentary Robot Condition). This finding speaks to the power of social presence in future human-robot interaction.

Was there a key factor that influenced whether participants kept Robovie's secret in the more stringent sense of the term (i.e., keeping the secret both times)? The answer is yes. Based on our binary logistic regression model, if participants believed that Robovie had low sociality (as measured by our social other scale), then it was essentially guaranteed that the participant did not keep the secret. When the sociality reached a threshold score of 5 or higher, participants had a roughly 50/50 chance of keeping the secret. Moreover, what was stunning to us was that the condition did not matter. Whether it was with Robovie in either of the two robot conditions or the human, the driving factor for the secret keeping was sociality, though sociality by itself did not guarantee it. Thus our results suggest that one's commitment to the sociality of a robot or human is a necessary but not sufficient condition for developing trust with that entity.

One implication of this result pertains to what can be called the "persuasive robot" literature, where robots are designed to encourage people to change their behavior. For example, Kidd and Breazeal [17] report on designing a robot that functions as a weight loss coach, and Belpaeme and colleagues [4] report on robot implementations in a hospital setting with children with diabetes, where one goal is to improve children's "management of their health condition" (p. 34). Based on our current results, it follows that for robots to be psychologically persuasive they need to elicit social attributions.

How, then, does one build that sociality into human-robot interaction? Our approach was to draw on and extend *Interaction Pattern Design*, described earlier in Methods. Namely, we instantiated—and used a good deal of space in this article to describe—more than a dozen high level interaction patterns, including *In Motion Together*, *Eliciting Personal Experience*, *Sharing Personal Interests & History*, *Collaboratively Generating Ideas to Solve a Mutually Recognized Problem*, *Expressing Emotion*, *Humor*, *Eliciting Moral Opinion*, *and Reflection on Shared Experience*. Based on our extensive micro-level behavioral coding, these interaction patterns proved effective in establishing the sociality of the human-robot interaction. Thus this study provides further evidence of a generalizable method and worthwhile design principles for establishing sociality in HRI.

Finally, it's worth asking a fundamental normative question. Should we even design and build our technological world such that we promote psychologically intimate relationships with robots [9, 11, 12]? One potential downside emerges from ideas that Dautenhahn [6] has raised in the context of HRI, where she draws on evolutionary theorists who have suggested that humans are limited in the number of people with whom they can form psychologically intimate relationships. The reasoning here is that for hundreds of thousands of years in our evolutionary history our numbers of psychologically intimate relations topped out at perhaps around 150 people, and likely were much fewer; and that those capabilities and limitations still form part of the architecture of the human mind. If so, then what if in the years to come many of these potential psychological "slots" are filled with intimacy not with people but with robots? Will that diminish the human experience, especially if those robots are diminished socially in other ways? Along similar lines, though from the standpoint of developmental psychology, there's the question of whether children growing up need a certain number of intimate and trusting human-human interactions, first with one's mother, as part of a healthy attachment bond [1, 5, 7], and then with other adults and also peers [21-22]. Is there a threshold for healthy child development that intimate trusting relationships with social robots can encroach upon, if not crossover? And, finally, taking a different tack, what if people develop intimate and trusting relationships with sex robots [18] and then find that those relationships with robots are easier to manage than with real people. Will that come at a cost to human development, human experience, and human flourishing? All of these questions, and more, become relevant based on the central finding of this study: that people's psychological intimacy with robots will increase as robots become increasingly social and pervasive in human lives.

5. ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. IIS-0905289. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. This work was also in part supported by the Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for Scientific Research Nos. 21110008 and 25240042, and by JSPS Core-to-Core Program, A. Advanced Research Networks. We thank Cady Stanton Laundry, Rose Maier, Nicole Jensen, and Nick McMillan for assistance with data collection and analysis, and Zakery Lee for technical assistance.

6. REFERENCES

[1] Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychology study of the strange situation*. Hillsdale, NJ: Lawrence Erlbaum.

[2] Baier, A. (1986). Trust and antitrust. *Ethics, 96*(2), 231-260. Retrieved from http://www.jstor.org/stable/2381376

[3] Bainbridge, W. A., Hart, J., Kim, E. S., & Scassellati, B. (2008). The effect of presence on human-robot interaction. *Proceedings of the 17th International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 701-706. doi:10.1109/ROMAN.2008.4600749

[4] Belpaeme, T., Baxter, P., Read, R., Wood, R., Cuayáhuitl, H., Kiefer, B., … Humbert, R. (2012). Multimodal child-robot interaction: Building social bonds. *Journal of Human-Robot Interaction, 1*, 33-53. doi:10.5898/JHRI.1.2.Belpaeme

[5] Bowlby, J. (1969). *Attachment and loss: Vol. 1. Attachment*. New York, NY: Basic Books.

[6] Dautenhahn, K. (2004). Socially intelligent agents in human primate culture. In R. Trappl & S. Payr (Eds.), *Agent culture: Human–agent interaction in a multicultural world* (pp. 45-71). Mahwah, NJ: Lawrence Erlbaum.

[7] Erikson, E. H. (1950). *Childhood and society*. New York, NY: Norton.

[8] Hoffman, G., & Vanunu, K. (2013). Effects of robotic companionship on music enjoyment and agent perception. *Proceedings of the 8th International Conference on Human-Robot Interaction*, 317-324. doi:10.1145/2447556.2447674

[9] Kahn, P. H., Jr. (2011). *Technological nature: Adaptation and the future of human life*. Cambridge, MA: MIT Press.

[10] Kahn, P. H., Jr., Freier, N. G., Kanda, T., Ishiguro, H., Ruckert, J. H., Severson, R., & Kane, S. K. (2008). Design patterns for sociality in human-robot interaction. *Proceedings of the 3rd International Conference on Human-Robot Interaction*, 97-104. doi:10.1145/1349822.1349836

[11] Kahn, P. H., Jr., Gary, H. E., & Shen S. (2013). Children's social relationship with current and near-future robots. *Child Development Perspectives*, 7, 32-37. doi:10.1111/cdep.12011

[12] Kahn, P. H., Jr., Ishiguro, H., Friedman, B., Kanda, T., Freier, N. G., Severson, R. L., & Miller, J. (2007). What is a human? Toward psychological benchmarks in the field of human-robot interaction. *Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems, 8*(3), 363-390. doi:10.1075/is.8.3.04kah

[13] Kahn, P. H., Jr., Kanda, T., Ishiguro, H., Freier, N. G., Severson, R. L., Gill, B. T., … Shen, S. (2012). "Robovie, you'll have to go into the closet now": Children's social and moral relationships with a humanoid robot. *Developmental Psychology, 48*, 303-314. doi:10.1037/a0027033

[14] Kahn, P. H., Jr., Kanda, T., Ishiguro, H., Gill, B. T., Ruckert, J. H., Shen, S., … Severson, R. L. (2012). Do people hold a humanoid robot morally accountable for the harm it causes? *Proceedings of the 7th International Conference on Human-Robot Interaction*, 33-40. doi:10.1145/2157689.2157696

[15] Kahn, P. H., Jr., & Turiel, E. (1988). Children's conceptions of trust in the context of social expectations. *Merrill-Palmer Quarterly, 34*, 403-419.

[16] Kaniarasu, P., & Steinfeld, A. M. (2014). Effects of blame on trust in human robot interaction. *Proceedings of the 23rd International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 850-855. doi:10.1109/ROMAN.2014.6926359

[17] Kidd, C. D., & Breazeal, C. (2007). A robotic weight loss coach. *Proceedings of the 22nd AAAI Conference on Artificial Intelligence*, 1985-1986. Retrieved from AAAI: http://www.aaai.org/Papers/AAAI/2007/AAAI07-366.pdf

[18] Levy, D. N. (2007). *Love + sex with robots: The evolution of human-robot relations*. New York, NY: HarperCollins.

[19] Nakagawa, K., Shiomi, M., Shinozawa, K., Matsumura, R., Ishiguro, H., & Hagita, N. (2011). Effect of robot's active touch on people's motivation. *Proceedings of the 6th International Conference on Human-Robot Interaction*, 465-472. doi:10.1145/1957656.1957819

[20] Nie, J., Pak, M., Marin, A. L., & Sundar, S. S. (2012). Can you hold my hand? Physical warmth in human-robot interaction. *Proceedings of the 7th International Conference on Human-Robot Interaction*, 201-202. doi:10.1145/2157689.2157755

[21] Rotenberg, K. J. (Ed.). (2010). *Interpersonal trust during childhood and adolescence*. New York, NY: Cambridge University Press.

[22] Rotter, J. B. (1980). Interpersonal trust, trustworthiness, and gullibility. *American Psychologist, 35*, 1-7. doi:10.1037/0003-066X.35.1.1

[23] Stanton, C. M., Kahn, P. H., Jr., Severson, R. L., Ruckert, J. H., & Gill, B. T. (2008). Robotic animals might aid in the social development of children with autism. *Proceedings of the 3rd International Conference on Human-Robot Interaction*, 271-278. doi:10.1145/1349822.1349858

[24] Turiel, E. (1998). The development of morality. In W. Damon (Series Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (pp. 863-932). New York, NY: Wiley.

[25] Velonaki, M., Silvera-Tawil, D., & Rye, D. (2013). Affective Human-Robot Interactions in Social Spaces: Two Case Studies. *Proceedings of the Workshop on Applications for Emotional Robots, 8th International Conference on Human-Robot Interaction*.

[26] Wada, K., & Shibata, T. (2007). Living with seal robots–Its socio psychological and physiological influences on the elderly at a care house. *IEEE Transactions on Robotics, 23*(5), 972-980. doi:10.1109/TRO.2007.906261

[27] Williams, V. S. L., Jones, L. V., & Tukey, J. W. (1999). Controlling error in multiple comparisons, with examples from state-to-state differences in educational achievement. *Journal of Educational and Behavioral Statistics, 24*, 42-69. doi:10.3102/10769986024001042