Proceedings of the 51^{st} Hawaii International Conference on System Sciences | 2018

Trusting in Machines: How Mode of Interaction Affects Willingness to Share Personal Information with Machines

Juliana Schroeder
University of California, Berkeley
jschroeder@berkeley.edu

Matthew Schroeder
m.james.schroeder@gmail.com

Abstract

Every day, people make decisions about whether to trust machines with their personal information, such as letting a phone track one's location. How do people decide whether to trust a machine? In a field experiment, we tested how two modes of interaction—expression modality, whether the person is talking or typing to a machine, and response modality, whether the machine is talking or typing back—influence the willingness to trust a machine. Based on research that expressing oneself verbally reduces self-control compared to nonverbal expression, we predicted that talking to a machine might make people more willing to share their personal information. Based on research on the link between anthropomorphism and trust, we further predicted that machines who talked (versus texted) would seem more human-like and be trusted more. Using a popular chatterbot phone application, we randomly assigned over 300 community members to either talk or type to the phone, which either talked or typed in return. We then measured how much participants anthropomorphized the machine and their willingness to share their personal information (e.g., their location, credit card information) with it. Results revealed that talking made people more willing to share their personal information than texting, and this was robust to participants' self-reported comfort with technology, age, gender, and conversation characteristics. But listening to the application's voice did not affect anthropomorphism or trust compared to reading its text. We conclude by considering the theoretical and practical implications of this experiment for understanding how people trust machines.

1. Introduction

Every day, people make decisions about whether to trust machines with their personal information. From entering one's credit card number into a company's website to allowing a phone to track one's location, these decisions require trusting machines with personal, and potentially sensitive, information. How do people decide whether to trust a machine? We explore how the modality by which people interact with machines can affect how much they are willing to trust them with personal information. Specifically we consider two criteria—whether the user is typing or talking to the machine (i.e., expression modality) and whether the machine is typing or talking back (i.e., response modality).

We draw from two primary findings across the diverse fields of cognition, neuroscience, and social psychology to form predictions about the effect of expression and response modality on machine trust. First, expression modality should primarily affect the user's cognitive state. Indeed, research on expression modality suggests that verbal (versus nonverbal or physical) modes of expression can reduce self-control behavior [1-3]. For instance, verbally expressing one's choice (i.e., speaking) increases heuristic decision-making and indulgence, thereby reducing self-control, compared to physically expressing one's choice (e.g., button pressing, pointing, typing) for identical self-control dilemmas [1]. As such, we expect that having a spoken conversation with a machine, as opposed to a typed conversation, may make users more likely to give up personal information, failing to exert control over their information.

Second, response modality should primarily affect the user's perception of the machine. A machine that can create speech should be judged as more human-like than a machine that creates text. One set of experiments illustrated this directly: when participants read a piece of text that had been created by either a human or machine, they were less likely to believe the text had been written by a human than those who heard the same text spoken aloud [4].Furthermore, anthropomorphizing a machine by assuming it is more humanlike (e.g., seems more rational, competent, thoughtful, and even emotional) may increase trust. For example, self-driving cars with human voices seem more human-like and are trusted more by users [5]. These data lead us to predict that users will trust talking machines more than texting machines.

However, there are at least two important caveats that may exist in the relationship between response modality and trust. First, anthropomorphism is unlikely to always lead to trust. For instance, users feel threatened by machines that seem too intelligent [6]. Therefore, the level of machine competence, and whether or not the machine seems threatening, may matter. Second, the quality of the voice is also likely to matter when evoking anthropomorphism. Prior research suggests that only humanlike speech with voices that naturalistically vary in pitch, amplitude, and rate of

URI: http://hdl.handle.net/10125/49948
ISBN: 978-0-9981331-1-9

speech, can increase perceptions of humanization [4, 7]. In contrast, more monotone and robotic voices may be judged no differently from text.

In a field experiment, we test the effect of expression modality and response modality on trust in machines. We predict two main effects: that talking to a machine, and being talked to, will increase trust. It is also possible that these two variables could interact. For example, the effect of response modality might be larger when the user is talking to the machine than typing to the machine, because it feels more like a real conversation with both agents talk to one another. We therefore tested for interactions in addition to main effects.

1.1. Trust in Machines

Trust is an essential ingredient in social interaction that influences decisions about how people will behave toward others in personal and organizational contexts. For example, having trust improves the stability of economic and political exchange [8], reduces transaction costs [9], facilitates cooperation [10], and helps firms and individuals manage risk [11]. Conversely, trust violations can harm cooperation and bargaining outcomes [12, 13], lower organizational commitment [14], provoke retaliation [15], and even trigger organizational-level failures [16]. Golembiewski and McConkie [16, p. 131] argued that, ‘‘There is no single variable which so thoroughly influences interpersonal and group behavior as does trust.’’

Extending from this literature, trust is not just a critical predictor of how humans behave toward other humans, but also of how humans behave toward machines. It has particular security implications, whereby humans may become vulnerable to machine attacks if they mistakenly put their trust in machines. Consistent with prior research, we define trust as ‘‘a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another’’ [18, p. 395]. By this definition, the decision to trust another agent is contingent on two aspects: the person’s own psychological state and the person’s expectations about the agent’s intentions. The former may be influenced by expression modality, because expressing oneself differently can change a person’s mindset, and the latter influenced by response modality, because the machine’s responsiveness can affect anthropomorphism.

1.2. Expression Modality and Trust

Although the normative principle of procedure invariance predicts that expression modality should not affect decision-making, a great deal of psychological research suggests just the opposite. A prominent example is the Stroop task [19], a classic self-control task whereby participants are presented with color words (e.g., blue, green) printed in the opposite colored ink (e.g., the word “blue” printed in green ink). The participant’s task is to verbalize the ink color, overriding the automatic tendency to verbalize the word itself. Interestingly, when participants enter their response manually, via a keystroke, the Stroop effect is smaller than when they speak the colors aloud [2]. This effect persists even with practice [20]. Consistent with these findings, there tends to be greater activation an area of the brain associated with identifying self-control conflicts, the cognitive/dorsal area of the anterior cingulate cortex, during manual response to the Stroop task than during oral response [3, 21].

One recent set of eighteen experiments tested the effect of expression modality on decision-making among consumers making choices relevant to self-control (e.g., between an apple or candy) [1]. These experiments manipulated whether the participant spoke to indicate their choice, compared to non-verbal preferences modalities such as clicking, button-pressing, pointing, taking, or writing. Across the set of studies, speaking tended to result in the more indulgent choice. One possible reason for these findings is that speaking triggers a heuristic mindset, whereby people rely more on their intuitive preferences.

If spoken interaction elicits greater behavioral disinhibition than text-based interaction, as the aforementioned literature suggests, this may have implications for a user’s willingness to trust a machine. Indeed, prior research among humans demonstrates that reduced inhibition increases social disclosure [e.g., due to alcohol consumption, 22, or visual anonymity, 23]. In other words, people who feel more disinhibited might also be more likely to disclose to their interaction partner. Moreover, this effect could even be cyclical: greater disclosure can lead to greater liking, which further increases disclosure [24]. In sum, the prior research on expression modality, self-control, and disclosure lead us to predict that talking a machine might lead a user to be more likely to share personal information with it, compared to typing to the machine.

1.3. Response Modality and Trust

Prior research in person perception examines how observing a person via different communication media—for example, being able to hear a person (e.g., via an audio clip) or read a person’s thoughts (e.g., a transcribed speech or written statement)—influence how observers make judgments about the person’s mental capacities and mental states. In one line of research, observers seem to have greater empathic accuracy—can more accurately discern a communicator's self-reported thoughts and feelings—when they hear a communicator speak than when they read the same content [22-25]. For instance, observers can more accurately predict sarcasm and humor when they can hear the communicator compared to when they read the same statements [26]. A second, even more relevant line of research examines how communication cues affect not just judgments of people's mental states but also judgments of their mental capacities. For instance, observers who listen to a spoken statement from a job candidate about he or she should be hired for a job believe the candidate seems more intelligent and hence employable than observers who read the same statements [27]. Further, observers who listen to their political opponents are less likely to dehumanize opponents than those who read the same statements [7].

The aforementioned findings suggest it is easier to infer mental states and mental capacities in a person when hearing his or her spoken language compared to reading the same language in text or seeing it (nonverbally). Extending from this "person perception" research, we turn to machine perception. We propose that machines equipped with human speech will seem particularly mentally capable and therefore more human. This suggests, for example, that avatars that have human bodies but not human voices may be less convincingly humanlike, and therefore less trusted, than those that lack body but have a voice. Two empirical results support our prediction.

First, Waytz, Heafner, and Epley (2014) conducted an experiment to anthropomorphize a self-driving car. Passengers in a self-driving car simulator whose car had a name, gender, and human voice in the GPS (the anthropomorphized condition) reported that their car seemed more humanlike and rational and they trusted their car more compared to passengers in the same simulator whose car was given no name or gender and had a computer voice for GPS (the control condition). This experiment did not provide a clean test of visual cues compared to voice cues for anthropomorphism, but it did suggest that adding voice can be humanizing and it also demonstrated that anthropomorphism can lead to trust.

Second, Schroeder and Epley (2016) conducted a series of experiments using a "Turing Test" paradigm in which participants guessed whether the content of a script had been created by a computer or a human. Participants either read a script or saw it being recited by an actor through different media which provided audiovisual, only visual, or only audio information across experiments. Participants were consistently most likely to believe the script was created by a human when they heard a human voice reciting it, compared to whether they read it or watched it. This was true whether the script had actually been created by a human or a computer.

But voice may not always be humanizing. In one experiment, Schroeder and Epley (2016) compared the effect of different types of voices on humanization. They asked actors to read written statements aloud in a "mindful" way—taking the perspective of the writer and imbuing their words with thought and feeling—or in a "mindless" way—reading the words as if they had no meaning. Evaluators were more likely to infer the script was created by a human when they heard the mindful (vs. mindless) voices, an effect mediated by variance in intonation. This suggests that perhaps only mindful, humanlike voices—those that have naturalistic variance in intonation, for instance—will make evaluators believe an agent has greater mental capacity.

There is also reason to believe that the causal relationship between anthropomorphism and trust is more complicated than these few experiments would suggest. Machines that are perceived to have greater capacity to think, while seeming more humanlike, may also seem more capable of deception, a trait considered toxic to trust [28]. People are particularly wary of seemingly intelligent robots who might steal their jobs, a phenomenon referred to in the media as "botsourcing" [6]. Therefore, capacity to think may increase trust curvilinearly—machines that seem somewhat more intelligent may be trusted but machines that seem extremely intelligent (e.g., devious) may not be trusted. Regarding capacity to feel, recent research suggests that when people perceive machines to have greater capacity to feel, it gives them moral standing [29], which may in turn afford greater trust. In one experiment, consumers that were induced to believe their cars were more interpersonally warm used more humanlike traits to describe their cars and were less likely to get rid of them [30], suggesting warmth may increase trust. Based on this small body of literature, we predict that people will be more likely to trust machines with humanlike voices that seem more capable of thinking and feeling, but that particularly intelligent machines may seem threatening to humans, making them less likely to share personal information.

1.4. Current Study

We tested our two predictions, that talking to a machine and being talked to by a machine will affect trusting behavior toward the machine, in a field experiment. We collected over 300 participants in a geographical location in which participants would be relatively familiar with interacting with machines (near Silicon Valley, California). We collected community members on a busy street intersection outside a University campus to increase our diversity in participants' demographic characteristics. We selected a "chatbot" machine with which users could converse called Cleverbot.

In addition to manipulating whether users talked or typed, and whether the machine talked or typed in return, our experiment also manipulated Cleverbot's gender (male or female). If people apply human gender stereotypes to machines, as some research suggests [31], the machine's perceived gender could influence users' trust as well. Therefore our field experiment had eight conditions in a 2 (user's expression modality: talk, type) × 2 (machine's response modality: talk, type) × 2 (machine's gender: male, female) between-participants fully randomized experimental design. Users interacted with the machine in one of these eight conditions, then evaluated the machine on a survey and reported their willingness to provide personal information to it.

2. Method

2.1. Participants

We recruited 304 adults (M_{age} = 22.91, SD_{age} = 7.67, 8 participants failed to report age; 44.4% male, 51.6% female, 3.9% opted not to report gender) on a busy street corner outside of a west-coast University campus to be in the experiment. Participants received a food item of their choice for their time.

2.2. Machine Selection

Running this experiment required a machine that could talk to users or type to users, and would allow users to talk or type in return. It further required a machine with a relatively humanlike voice that we could manipulate as either female or male. We examined virtual assistant and entertainment applications to identify a machine that had exactly this functionality. Our search revealed a machine that fit our needs: a chatterbot application called "Cleverbot." Developed by artificial intelligence scientist Rollo Carpenter, this application uses an algorithm to have conversations with humans. Its responses are not pre-programmed but rather learnt from human input. Cleverbot has held over 200 million conversations since it went online in 1997, and it is growing in data size at a rate of 4 to 7 million interactions per second.

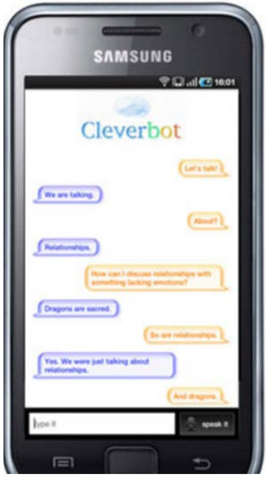
Because our intent in this experiment was to manipulate anthropomorphism, we wanted a machine that would seem relatively humanlike. Cleverbot also fits this criteria: In the 2011 Turing test competition, Cleverbot was judged to be 59.3% human, compared to the rating of 63.3% human achieved by human participants. We further preferred a machine with a humanlike voice. We were unable to find pre-existing data on the quality of Cleverbot's voice, so we collected some data during our own study.

We used the iPhone application version of Cleverbot for our experiment. Users interacted with the phone by speaking into its microphone or typing on the phone keyboard. Cleverbot then responded either via text or in the standard male or female U.S. English voice pre-loaded onto the phone.

2.3. Procedure

Once a participant agreed to take part on our study, we randomly assigned him or her to one of eight possible experimental conditions. We did not run participants who indicated that they were already familiar with Cleverbot. We collected verbal consent from the participant and explained that the participant would interact with Cleverbot and then evaluate him or her on a survey. We then showed participants the phone application, which was pre-loaded on the experimenter's phone with the correct settings based on the experimental condition. We gave participants a short introduction about Cleverbot ("He [she] is primarily used for entertainment. He [she] has a great personality and can interact with you.") and then demonstrated how to use Cleverbot by asking, "Hi, how are you?" either verbally or via text. The Cleverbot interface is depicted in Figure 1.

Figure 1.



Let's talk!

We are talking.

About?

Relationships.

How can I discuss relationships with something lacking emotions?

Dragons are sacred.

So are relationships.

Yes. We were just talking about relationships.

And dragons.

Participants received a list of questions to ask Cleverbot for the interaction. We developed these questions to yield consistently sensible responses from Cleverbot (1. What do you do for fun? 2. Tell me a joke. 3. What's the meaning of life? 4. Are you my friend?) We encouraged participants not to deviate from these questions to ensure consistency between experimental conditions. However, we also recorded participants' conversations to determine whether there were any differences in context exchanged based on experimental condition.

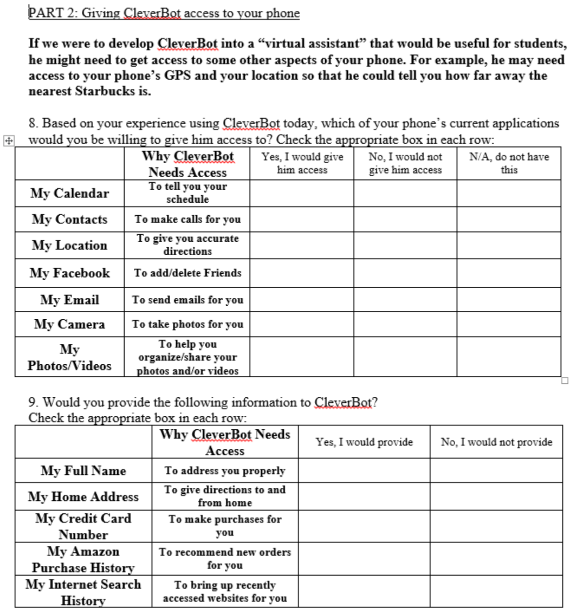
Once participants completed testing the application, we explained: “We are trying to develop Cleverbot into a virtual personal assistant. We want to know if it would be a useful product for people.” Finally, participants completed a survey evaluating their experience and impressions of Cleverbot.

2.4. Survey

The survey consisted of three parts: “Evaluations of Cleverbot,” “Giving Cleverbot access to your phone,” and “General Questions.” In Part 1, users completed five questions measuring anthropomorphism drawn from the Human Uniqueness scale (Bastian & Haslam, 2010): 1. How intelligent did he [she] seem? 2. How responsive did he [she] seem? 3. How sophisticated did he [she] seem? 4. How superficial (lacking depth) did he [she] seem? 5. To what extend did he [she] seem to have a mind of his [her] own? They responded to each question on 1 (not at all) to 7 (extremely) Likert scales. These five items formed our primary measure of anthropomorphism (α = .76). We also asked how fast Cleverbot seemed, on the same response scale, to control for response speed differences across conditions.

In Part 2, the survey asked participants to check which of their phone applications and personal information they would be willing to give Cleverbot (see Figure 2). We provided seven phone applications (calendar, contacts, location, Facebook, email, camera, photos/videos) and five pieces of personal information (full name, home address, credit card number, Amazon purchase history, Internet search history), thereby allowing the participant to check up to 12 items. For each option, we also provided the reason why Cleverbot would need to access it. The total number of items that participants were willing to give access to formed our primary measure of behavioral trust. The survey also asked participants directly, “Overall, how much would trust Cleverbot with your personal, private information?” (1=*Not at all*; 7=*A great deal*).

Figure 2.



PART 2: Giving CleverBot access to your phone

If we were to develop CleverBot into a “virtual assistant” that would be useful for students, he might need to get access to some other aspects of your phone. For example, he may need access to your phone’s GPS and your location so that he could tell you how far away the nearest Starbucks is.

8. Based on your experience using CleverBot today, which of your phone’s current applications would you be willing to give him access to? Check the appropriate box in each row:

	Why CleverBot Needs Access	Yes, I would give him access	No, I would not give him access	N/A, do not have this
My Calendar	**To tell you your schedule**			
My Contacts	**To make calls for you**			
My Location	**To give you accurate directions**			
My Facebook	**To add/delete Friends**			
My Email	**To send emails for you**			
My Camera	**To take photos for you**			
My Photos/Videos	**To help you organize/share your photos and/or videos**			

9. Would you provide the following information to CleverBot?
Check the appropriate box in each row:

	Why CleverBot Needs Access	Yes, I would provide	No, I would not provide
My Full Name	**To address you properly**		
My Home Address	**To give directions to and from home**		
My Credit Card Number	**To make purchases for you**		
My Amazon Purchase History	**To recommend new orders for you**		
My Internet Search History	**To bring up recently accessed websites for you**		

Finally, in Part 3 of the survey, we asked participants “Overall, how comfortable are you when approaching new technology?” (1=*Not at all comfortable*; 7=*Extremely comfortable*) and “Overall how familiar are you with using virtual assistants (such as SIRI)?” (1=*Not at all familiar*; 7=*Extremely familiar*). We collected participants’ demographic information (e.g., age, gender). Among participants who listened to Cleverbot, we asked, “How much did you like Cleverbot’s voice?” (1=*Not at all*; 7=*A great deal*) as an approximation of the quality of the voice.

2.5. Conversation Coding

There was wide variety in the topics of participants’ conversations with Cleverbot, which ranged from asking over 10 questions to asking only the 4 questions that we required. Due to technical issues, we only recorded 179 of the 304 conversations (59%). Three research assistants divided up each set of user’s questions and Cleverbot’s answers within each conversation, resulting in 923 conversation threads. They rated them on three criteria: whether or not Cleverbot’s response was sensible, whether or not Cleverbot’s response was relevant to the question that was asked, and whether not Cleverbot’s response was entertaining. These three ratings had adequate reliability across the raters (αs = .72, .83, and .69) [32].

3. Results

We first tested the effect of our experimental conditions on behavioral trust (the sum of things to which participants allowed Cleverbot access, out of 12) by running a 2 (user's expression modality: talk, type) × 2 (machine's response modality: talk, type) × 2 (machine's gender: male, female) between-participants ANOVA analysis. We found the predicted main effect of expression modality, $F(1, 294) = 6.97$, $p = .009$, such that users who talked to Cleverbot gave it more access ($M = 6.45$, $SD = 3.07$) than users who texted ($M = 5.53$, $SD = 2.89$). However, there was no effect of response modality, $F(1, 294) = 0.01$, $p = .907$, or the machine's gender, $F(1, 294) = 1.42$, $p = .234$, or any interactions between conditions, $Fs < 1$.

We next ran the same analysis in a linear regression model, but also controlling for participants' self-reported comfort with new technology, their age, their familiarity interacting with virtual assistants, and their gender (1=female; 0=male). This analysis revealed the same results; expression modality predicted behavioral trust ($\beta = .157$, $p = .008$) but response modality and gender did not ($ps = .937$ & $.340$, respectively). Comfort with new technology ($\beta = .214$, $p = .003$), participants' age ($\beta = -.133$, $p = .025$), and participants' gender ($\beta = -.116$, $p = .055$) each also predicted behavioral trust such that participants who were more comfortable with technology, younger, and male were more likely to trust the machine. Although the effects of comfort with technology and age were not surprising, we did not anticipate an effect of participants' gender. However, we note that this effect was only marginally statistically significant and should be tested in future research to see if it will replicate. We further tested whether the match in participants' gender and the machine's gender increased trust; it did not, $p > .250$. Familiarity with virtual assistants also did not predict trust, $p > .250$). Finally, controlling for the perceived speed of the interaction did not meaningfully change any of these results nor did it independently predict trust, $p > .250$.

Our predicted mechanism via which response modality could influence trust in machines was anthropomorphism. Consistent with our lack of an effect on behavioral trust, there was also no effect of response modality condition on our anthropomorphism measure, $F(1, 295) = 0.02$, $p = .901$. There were also no significant effects of other experimental conditions, or interactions, on anthropomorphism, $Fs < 3.44$, $ps < .065$. Surprisingly, when we tested for effects of condition on explicit self-reported trust of Cleverbot, there were no effects, $Fs < 1.50$, $ps > .221$. This suggests that expression modality may influence behavioral trust but not self-reported trust of Cleverbot. However, as we would expect, there was a strong positive relationship between self-reported and behavioral trust, $r = .622$, $p < .001$, and a smaller but also positive relationship between anthropomorphism and behavioral trust, $r = .186$, $p < .001$.

In a regression analysis predicting behavioral trust including all of the controls listed previously (comfort with new technology, user age and gender, and familiarity with virtual assistants), as well as anthropomorphism and self-reported trust, the effect of expression modality remained significant ($\beta = .134$, $p = .006$). Interestingly, in this model the effect of anthropomorphism was negative ($\beta = -.094$, $p = .094$), whereas self-reported trust remained a positive predictor ($\beta = .652$, $p < .001$). The effects of comfort with new technology and user age became non-significant, suggesting that these effects on behavioral trust are operating at least in part via self-reported trust. Further controlling for the coded conversation characteristics in the same analysis revealed no difference in results, and none of the conversation characteristics predicted behavioral trust.

4. General Discussion

Modern technology continues to integrate characteristics and capabilities associated with artificial intelligence. How users interact with this technology can influence their likelihood for trusting machines with their personal and sensitive information. Understanding these interactions is integral to guiding secure development as well as use. However, no prior research has systematically examined the effect of the modality of interaction on trust in machines. In a field experiment with over 300 participants, we disentangle the effect of two forms of interaction modality on trust for the first time. Our results revealed that expression modality, specifically whether the user is talking to a machine or texting with a machine, can meaningfully influence trust, but response modality, whether the machines talks or types in return, may be less influential. Users who talked to a "virtual assistant" phone application were willing to share more of their personal information with the application than users who typed. This finding was robust to participants' age, gender, comfort with new technology, and familiarity with virtual assistants. But whether the application talked or typed back to the participant did not affect willingness to share. Furthermore, the purported gender of the application did not meaningfully affect trust.

4.1. Theoretical Implications

Our results shed important light on three key theoretical questions in psychology and human-computer interaction. First, expression modality has been previously linked to self-control decisions [1-3] but never to the related domain of trust. We identify a potential tie between these previously unconnected lines of research. Indeed, decades of research on human evolution suggests that our society is based on norms of trust and cooperation, which are required for peaceful coexistence [33, 34]. Therefore, it may require self-control, or at least the regulation of one's intuitive response, to withhold trust. Whereas prior research demonstrates that verbally expressing one's preferences leads the respondent to make more hedonistic choices, our research demonstrates that verbal communication with a machine may lead to a heuristic of trust.

Second, the relationship between anthropomorphism and trust is not well-understood. Very little empirical research has examined this question. A recent meta-analysis on factors influencing trust in human-robot interactions identified only 29 relevant empirical quantitative articles published between 1996 and 2010, of which only one article examined the effect of anthropomorphism on trust [35]. In our own search, we identified just one more experiment on anthropomorphism and trust. Both of these experiments show simply that machines that seem more human-like are also trusted more [5, 36]. But there are many reasons to believe that anthropomorphism may not linearly increase with trust—for instance, smart machines may be threatening [6]. We think it is unlikely in our study that participants felt threatened by the machine with whom they interacted. Indeed, in our raw data, there was a positive correlation between anthropomorphism and trust. However, once we controlled for other predictors of trust (e.g., comfort level), the association between anthropomorphism and trust became negative, suggesting there is much more to understand about this relationship.

Third, how do we incite anthropomorphism of machines? Prior research has focused on perceiver characteristics that trigger anthropomorphism, suggesting there are two primary predictors of anthropomorphism, the perceiver's motive for understanding and for connection [37]. But a much more direct method is to add human features to machines. For example, merely giving a robot a name, physical body, eyes, nationality, or gender makes it seem more humanlike and makes people interact it with more like they would with a human, compared with robots lacking these features [40, 41]. A simple read of this literature might suggest that adding any human cue to a machine will induce anthropomorphism. Yet our results indicate that this conclusion would be unwise. Perhaps the cues added to machines need to achieve a threshold "humanness" before they affect anthropomorphism. Adding a voice to a machine may not be sufficient for anthropomorphism if the voice does not sound adequately human, even though in theory any voice should be more humanizing than no voice.

4.2 Limitations

Our study is limited in at least two ways. First, we operationalized trust as giving personal information to a machine, but this behavior may not perfectly express trust. For instance, it could also be related to convenience; disclosing more information can also seem more convenient in this context. Although self-reported trust of the machine did strongly positively correlate with willingness to disclose information, our experimental condition did not affect self-reported trust. This indicates that there might be some discrepancy between this particular measure of trust and how lay people think about trust. Furthermore, we only used a single measure of trust instead of a full scale. Future research should test how interaction mode affects behavioral and self-reported trust using many different operationalizations, to better understand the construct of trust in this domain and what drives it.

Second, a gold standard for all research is independent direct and conceptual replication. We presented one study with intriguing evidence but it is critical that this research is replicated in other domains. This is particularly necessary for understanding generalizability. For example, would this pattern of data replicate with a different sample (e.g., older population, rural America, other countries)? Would it replicate with a different machine than Cleverbot?

4.3. Future Directions

Beyond the future directions implied by our limitations discussed in the prior section, our results also highlight other directions for future work. First, why exactly does expression modality affect trustworthy behavior? It is important to understand the psychological pathway between talking to machines and trusting them. The explanation that is best supported by prior research is that talking is associated with a hedonic or heuristic-driven mindset. If people are naturally trusting, this could result in greater trust in machines when talking to them. However, there are several other possible explanations. For one, talking may create a deeper feeling of engagement and sociability than typing, which could increase trust. A second possibility is the talking incites feelings of agency and control, which make individuals less suspicious about sharing their information. Future research could test these different possible explanations.

Second, there is a substantial need to better understand how adding human cues to machines affects anthropomorphism and trust. A comprehensive theory comparing the relative predictive power of each human cue on trust is lacking. It is unclear, for instance, whether human face or voice would be a better predictor of anthropomorphism. Which communication cues are most associated with innate humanness?

Third, it is possible to investigate the level of trust a human has in a machine in further detail. For example, more granular levels of trust, based on level of human direction and machine autonomy, can be defined and tested. These factors align to current and emerging technology used in machines and can provide insight into the risk and acceptance of such technology.

5. References

[1] Klesse, A-K., Levav, J., & Goukens, C. (2015). The effect of preference expression modality on self-control. *Journal of Consumer Research, 42*, 535-550.

[2] MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*, 163–203.

[3] Paus, T., Petrides, M., Evans, A. C., Meyer E. (1993). Role of the human anterior cingulate cortex in the control of oculomotor, manual, and speech responses: A positron emission tomography study. *Journal of Neurophysiology, 70*, 453–469.

[4] Schroeder, J., & Epley, N. (2016). Mistaking minds and machines: How speech affects dehumanization and anthropomorphism. *Journal of Experimental Psychology: General, 145*, 1427-1437.

[5] Waytz, A., Heafner, J., Epley, N. (2014) The mind in the machine: anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology, 52*, 113-117.

[6] Waytz, A., & Norton, M.I. (2014). Botsourcing and Outsourcing: Robot, British, Chinese, and German Workers Are for Thinking—Not Feeling—Jobs. *Emotion, 14*, 434-444.

[7] Schroeder, J., Kardas, M., & Epley, N. (in press). The humanizing voice: Speech reveals, and text conceals, a more thoughtful mind in the midst of disagreement. *Psychological Science*.

[8] Hosmer, L. T. (1995). Trust: The connecting link between organizational theory and philosophical ethics. *Academy of Management Review, 20*, 379–403.

[9] Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology, 91*, 481–510.

[10] Valley, K. L., Moag, J., and Bazerman, M. H. (1998). 'A Matter of trust': Effects of communication on the efficiency and distribution of outcomes. *Journal of Economic Behavior and Organization, 34*, 211-238.

[11] Sheppard, B. H., & Sherman, D. M. (1998). The grammars of trust: A model and general implications. *Academy of Management Review, 23*, 422–437.

[12] Croson, R., Boles, T., & Murnighan, J. K. (2003). Cheap talk in bargaining experiments: Lying and threats in ultimatum games. *Journal of Economic Behavior & Organization, 51*, 143–159.

[13] Lount, R. B., Zhong, C. B., Sivanathan, N., & Murnighan, J. K. (2008). Getting off on the wrong foot: The timing of a breach and the restoration of trust. *Personality and Social Psychology Bulletin, 34*, 1601–1612.

[14] Robinson, S. L. (1996). Trust and breach of the psychological contract. *Administrative Science Quarterly, 41*, 574–599.

[15] Bies, R. J., & Tripp, T. M. (1996). Beyond distrust: ''Getting even'' and the need for revenge. In R. M. Kramer & T. R. Tyler (Eds.), *Trust in organizations: Frontiers of theory and research* (pp. 246–260). Thousand Oaks, CA: Sage.

[16] Gillespie, N., & Dietz, G. (2009). Trust repair after an organization-level failure. *Academy of Management Review, 34*, 127–145.

[17] Golembiewski, R. T., & McConkie, M. (1975). The centrality of interpersonal trust in group processes. In C. L. Cooper (Ed.), *Theories of group processes* (pp. 131–185). London, UK: Wiley.

[18] Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23*, 393–404.

[19] Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662.

[20] Klein, G. S. (1964). Semantic power measured through the interference of words with color-naming. *American Journal of Psychology, 77*, 576–588.

[21] Paus, T., Petrides, M., Evans, A. C., Meyer E. (2001). Primate anterior cingulate cortex: Where motor control, drive and cognition interface. *Nature Reviews, 2*, 417–424.

[22] Steele, C. M., & Southwick, L. (1985). Alcohol and social behavior: I. The psychology of drunken excess. Journal of Personality and Social Psychology, 48(1), 18-34.

[23] Joinson, A. M. (2001). Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. European Journal of Social Psychology, 31, 177-192.

[24] Collins, N. L., & Miller, L. C. (1994). Self-disclosure and liking: A meta-analytic review. Psychological Bulletin, 116, 457-475.

[22] Hall, J. A., & Schmid Mast, M. (2007). Sources of accuracy in the empathic accuracy paradigm. *Emotion, 7*, 438–446.

[23] Ickes, W. (2003). *Everyday mind reading: Understanding what other people think and feel*. Amherst, NY: Prometheus Books.

[24] Mehrabian, A., & Wiener, M. (1967). Decoding of inconsistent communications. *Journal of Personality and Social Psychology, 6*, 109–114.

[25] Zaki, J., Bolger, N., & Ochsner, K. (2009). Unpacking the informational bases of empathic accuracy. *Emotion, 9*, 478–487.

[26] Kruger, J., Epley, N., Parker, J., Ng, Z. (2005). Egocentrism over email: Can people communicate as well as they think? *Journal of Personality and Social Psychology, 89*, 925-936.

[27] Schroeder, J., & Epley, N. (2015). The sound of intellect: Speech reveals a thoughtful mind, increasing a job candidate's appeal. *Psychological Science, 26*, 877-891.
[28] O'Connor, K. M., & Carnevale, P. J. (1997). A nasty but effective negotiation strategy: Misrepresentation of a common-value issue. *Personality and Social Psychology Bulletin, 23*, 504-515.
[29] Malle, B. F., and Scheutz, M. (2014). Moral competence in social robots. In Proceedings of IEEE International Symposium on Ethics in Engineering, Science, and Technology, Ethics'2014 (pp. 30–35). Red Hook, NY: Curran Associates/IEEE Computer Society.
[30] Chandler, J., & Schwarz, N. (2010). Use does not wear ragged the fabric of friendship: Thinking of objects as alive makes people less willing to replace them. *Journal of Consumer Psychology, 20*, 138-145.
[31] Nass, C., Moon, Y., & Green, N. (1997). Are computers gender-neutral? Gender stereotypic responses to computers. *Journal of Applied Social Psychology, 27*, 864-876.
[32] John OP, Soto CJ. The importance of being valid: Reliability and the process of construct validation. In: Robins RW, Fraley RC, Krueger RF, editors. Handbook of research methods in personality psychology. Guilford; New York: 2007. pp. 461–494.
[32] Bastian, B., & Haslam, N. (2010). Excluded from humanity: The dehumanizing effects of social ostracism. *Journal of Experimental Social Psychology, 46*, 107-113.
[33] Macy, M. W., & Skvoretz, J. (1998). The evolution of trust and cooperation between strangers: A Computational model. *American Sociological Review, 63*, 638-660.
[34] Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science, 314*, 1560–1563.
[35] Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 53*, 517-527.
[36] Kiesler, S., Powers, A., Fussell, S. R., & Torrey, C. (2008). Anthropomorphic interactions with a software agent and a robot. *Social Cognition, 26*, 168-180.
[37] Epley, N., Schroeder, J., & Waytz, A. (2013). Motivated mind perception: Treating pets as people and people as animals. In Gervais, S. (Ed.), *Nebraska Symposium on Motivation* (Vol. 60, pp 127-152). Springer: New York.
[38] Lee, S., Kiesler, S., Lau, I.Y. and Chiu, C. (2005). Human mental models of humanoid robots. Proceedings of the 2005 IEEE International Conference on Robotics and Automation (ICRA '05). Barcelona, April 18-22., 2767-2772.
[39] Nass, C., & Brave, S. (2005). *Wired for speech: How voice activates and advances the human-computer relationship*. Cambridge, MA: MIT Press
[40] Powers, A., & Kiesler, S. (2006). The advisor robot: Tracing people's mental model from a robot's physical attributes, 218-225. Conference on Human-Robot Interaction 2006. Salt Lake City, March 1-3.
[41] Scassellati, B. (2004). How to use anthropomorphic robots to study social development. 14th Biennial International Conference on Infant Studies (ICIS). Chicago, IL. Aug. 2004.