Machine Heuristic: When We Trust Computers More than Humans with Our Personal Information

S. Shyam Sundar
Media Effects Research Laboratory
Penn State University
University Park, PA, USA
sss12@psu.edu

Jinyoung Kim
Media Effects Research Laboratory
Penn State University
University Park, PA, USA
juk3151120@gmail.com

ABSTRACT

In this day and age of identity theft, are we likely to trust machines more than humans for handling our personal information? We answer this question by invoking the concept of "machine heuristic," which is a rule of thumb that machines are more secure and trustworthy than humans. In an experiment (N = 160) that involved making airline reservations, users were more likely to reveal their credit card information to a machine agent than a human agent. We demonstrate that cues on the interface trigger the machine heuristic by showing that those with higher cognitive accessibility of the heuristic (i.e., stronger prior belief in the rule of thumb) were more likely than those with lower accessibility to disclose to a machine, but they did not differ in their disclosure to a human. These findings have implications for design of interface cues conveying machine vs. human sources of our online interactions.

CCS CONCEPTS

• Human-centered interaction (HCI)~Empirical studies in HCI
• Human and societal aspects of security and privacy

KEYWORDS

Virtual Agent, Cognitive Heuristics, Machine Heuristic, Automation Bias, MAIN Model, Secure and Trustworthy Computing

ACM Reference format:
S. Shyam Sundar and Jinyoung Kim. 2019. Machine Heuristic: When We Trust Computers More Than Humans With Our Personal Information. In *2019 CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4-9, Glasgow, Scotland, UK.* ACM, New York, NY, USA. Paper 538, 9 pages. https://doi.org/10.1145/3290605.3300768

CHI 2019, May 4-9, 2019, Glasgow, Scotland, UK.

ACM ISBN 978-1-4503-5970-2/19/05...$15.00.
DOI: https://doi.org/10.1145/3290605.3300768

1 INTRODUCTION

Online and mobile users conduct a variety of transactions, from shopping to banking to even counseling, by chatting with conversational agents. Increasingly, these agents are fully automated computer systems that function autonomously without requiring human operators. Studies have shown that users are quite willing to reveal personal information to such chatbots [16]; in fact, users are more honest and feel more comfortable revealing sensitive information to virtual agents than to humans because these agents would not judge them [22].

While the process of revealing information to a virtual agent may pose less of a "face threat" [10] and therefore lower impression-management concerns among users, it does not guarantee the security of that information. Personal data are often tracked, shared, and even sold to third parties without the user's explicit permission [1]. In fact, the terms and conditions outlined in the privacy policy of most services and applications explicitly acknowledge this possibility. For example, Snapchat's updated privacy policy implies that the app implements computer algorithms that collect almost all information about users, including their name, exact location, friends, and messages, and save the data for a period of time even after their account is deleted [14]. This widespread practice should raise grave concerns from users about privacy breaches of data provided to networked systems of any kind. So, while it may be true that machines do not judge users in the same manner as some humans, they tend to store user data that may later be used to judge them. Similarly, while it seems intuitive that machine agents do not gossip about users in the same manner as some human agents, they in effect perform a similar function by sharing user information with different servers and systems.

However, users seldom think about such risks when disclosing their personal information to machine agents. They tend to be "privacy pragmatists" with an optimistic outlook [30]. More generally, the apprehension about privacy breaches among many users does not appear to manifest in their everyday online behaviors. For example, they mindlessly consent to various privacy terms and conditions and share their contact information and personal photos/videos to mobile applications without thinking about the consequences [20]. That is, they behave in a way that contradicts their conservative attitudes toward privacy. This inconsistency between users' conservative attitudes and their actual online behaviors is known in the literature as "privacy paradox" [19].

Why do users reveal more personal information even if they say they are very concerned about their privacy? Scholars from the social cognition tradition explain that users are "cognitive misers" [9] who make disclosure decisions on the basis of cognitive heuristics or mental shortcuts (i.e., rules of thumb) [29], often triggered by contextual cues in the situation [13] or on the interface [25], rather than going through careful and effortful analysis of risks and benefits involved in each transaction [e.g., 26, 32]. One heuristic that is pertinent to users' instinctive preference for machine agents over human agents is the *machine* heuristic, the rule of thumb that machines are more objective than humans, can perform tasks with greater precision and handle information in a more secure manner. This means interface cues signaling that the chat agent is fully automated are likely to engender greater trust among users, especially those who believe strongly in this heuristic. We test this possibility in the current study, as described in the following sections.

2 ROLE OF COGNITIVE HEURISTICS IN PRIVACY DECISION-MAKING

As many online transactions demand a substantial amount of cognitive energy to process information, careful assessment of the security of various interfaces is not always feasible. Specifically, various contextual factors underlying online and mobile contexts, such as users' time constraints, lack of technological efficacy [7], and information overload [23], make it difficult to comply with the privacy protection ideals that users want to attain. Therefore, users tend to make decisions that maximize efficiency at the cost of thoroughness [9]. In fact, Sundar and his team [26] demonstrated that interface cues on e-commerce websites triggered relevant cognitive heuristics, which helped minimize the cognitive resources necessary for analysis of each online transaction and thereby promoted expedient decision-making. In making this claim, they relied on the Modality-Agency-Interactivity-Navigability (MAIN) model [25], which proposes that technological affordances (in the form of interface cues) elicit specific cognitive heuristics that help users make a quick decision about the credibility of a given information system and its content. For example, the provision of other shoppers' opinions about a product in the form of interface cues such as star ratings can trigger the bandwagon heuristic and thereby sway users' opinions in the direction of the bandwagon, without the user perusing the product information on the site.

Drawing on this theoretical framework, Sundar et al. [26] examined if online shoppers revealed more personal information when they were notified of the potential benefits of information sharing (e.g., the personalized shopping services), which triggers the *benefit* heuristic (i.e., my personal information will be used to benefit me). Also, they investigated whether users disclosed less information if they were informed of the risk of privacy breaches due to the operation of the *fuzzy boundary* heuristic (i.e., my information might be shared with third parties, therefore it is unsafe to reveal it). Results showed that users who were primed with the fuzzy boundary heuristic showed lesser intentions for information disclosure, whereas those cued with the benefit heuristic wanted to reveal more information. In a similar vein, Zhang and her colleagues [32] tested whether a mobile security-warning message lowers users' perceived trust and positive attitudes toward the mobile interface due to the operation of the *online security* heuristic (i.e., online is not safe, thus it is risky to reveal my information). They showed that participants were less likely to reveal their information to the site since the warning message primed the negative cognitive heuristic in the minds of users.

3 MACHINE HEURISTIC

One of the key heuristics identified by the MAIN model [25] is the "machine heuristic." It refers to the mental shortcut wherein we attribute machine characteristics or machine-like operation when making judgments about the outcome of an interaction. This heuristic is said to be triggered when cues on the interface suggest that the user is dealing with a machine rather than a human. The model posits that machine agency is made increasingly possible by "agency affordances" of modern computer systems which allow for the possibility that not only programmers and other users, but also the computer systems themselves can serve as agents of interaction. When the perceived locus of our interaction is a machine, rather than another human being, the model states that we automatically apply common stereotypes about machines, namely that they are mechanical, objective, ideologically unbiased and so on. Such perceptions of machine operations may not always be accurate—machines do fail and they can be prone to error and manipulation. In recent times, computer systems are seen as being increasingly vulnerable to hacking, and algorithms are increasingly seen in the computing community as being driven by the agenda of their creators and sometimes even users. But, these elaborated notions of machine weaknesses may not be at the forefront of a user's mind when interacting online. Instead, the most immediately accessible ideas about machines are the ones that are well-entrenched in our mental models, which tend to be the positive stereotypes about machine infallibility and neutrality.

Studies have shown that users believe that machines are less likely to misuse sensitive information because their performance is generally associated with precision and low error rates [e.g., 3, 28]. In the recommender systems literature, studies have shown that humans tend to attribute greater power and trust to machines and their recommendations compared to other sources of recommendation—a phenomenon called "automation bias" [18]. Humans have a tendency to favor directives from automated decision aids or decision support systems and follow the recommendations of the aids even if it is incorrect or suboptimal, while ignoring the recommendations made without automation [e.g., 15]. Scholars have found evidence of automation bias in various domains, including aviation, healthcare, process control, and command-and-control operations in military contexts, demonstrating how decision-making can be biased due to users' over-reliance on automated aids and decision support systems.

Such ingrained faith in automated operations underlies the "machine heuristic," which is easily triggered when a machine, rather than a human, is the attributed source or locus of one's interactions. When applied to the context of privacy decision-making, this heuristic could result in greater compliance with system request. That is, users' tendency for disclosing personal information is likely to be higher when there are interface cues that represent automated decision aids (e.g., a virtual assistant like Siri or Alexa) that request such information rather than a human being at the other end (e.g., a customer-service agent), because the automated aids trigger the *machine* heuristic and thereby provide them a higher sense of security. Based on this rationale, we propose the following hypothesis for study:

H1: If an interface provides cues suggesting that a machine, rather than a human, is handling their information, users are more likely to reveal personal information.

Empirical support for this hypothesis will not by itself demonstrate that users applied the mental shortcut of "machine heuristic" in making their decision to disclose personal information. For a heuristic to be operational, it has to be cognitively accessible at the time of decision-making [11]. That is, the rule of thumb (that machines are superior and more trustworthy) ought to be foremost in the minds of users when they are deciding whether or not to disclose their personal information. One way to ensure this is by priming users about the heuristic just before their exposure to the stimulus in the study [5], but this is not feasible in a normal user context. So, we adopted a different strategy, by measuring individual differences in their trait belief in machine heuristic, with the logic being that for those with stronger belief, the heuristic would be chronically accessible and therefore readily applied at the time of decision-making. If machine heuristic is indeed operational in H1, then we should find that those with a stronger belief in the heuristic would more strongly favor machine agents over human agents. This leads us to an interaction hypothesis, wherein trait belief in machine heuristic is said to moderate the effect hypothesized in H1. Formally stated,

H2: The greater the belief in machine heuristic, the higher the disclosure of information to a machine agent, but not to a human agent.

That is, if an interface provides cues suggesting that a *machine* is handling their information, users who have greater belief in the machine heuristic are more likely to reveal personal information than users who have lesser belief in the machine heuristic, but this difference will not be found when the interface suggests that the user is interacting with a human agent.

4 METHOD

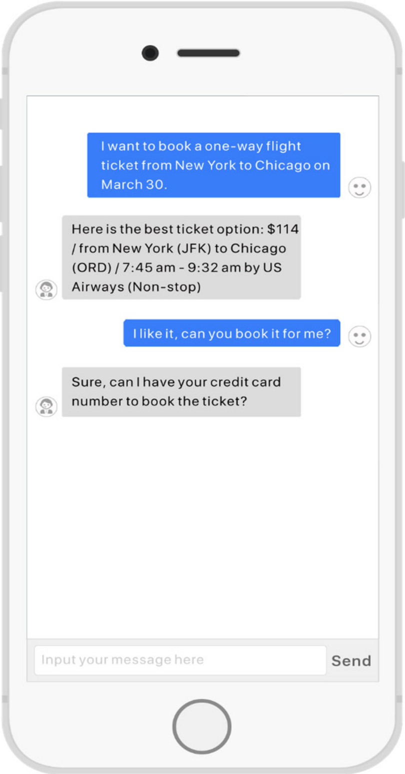
A scenario-based, between-subjects experiment was conducted to test the proposition that the machine heuristic will affect users' intention to disclose private information, and that this association is conditional upon users' degree of belief in the heuristic. Users were exposed to an online chat with a human or a machine chat agent that searched for a flight ticket between two cities in the US and then asked to report their intention to share their credit card number with the chat agent in order to purchase the ticket.

4.1 Participants

One hundred and sixty participants were recruited from Amazon Mechanical Turk (MTurk). All participants resided in the US and had a Human Intelligence Task (HIT) approval rate over 90% (i.e., at least 90% of previous HITs completed by them had been approved, rather than rejected, by their work requesters). Among the participants, 86 were males (53.8%) and 74 were females (46.3%), with the average age of 38.29 (*SD* = 12.51, range = 18-75). The majority of the sample was Caucasian (71.9%), with a wide range of educational background, from high-school to graduate degree. The plurality of participants reported having a bachelor's degree (41.9%). Eighty-four participants were randomly assigned to the human-agent condition, while 76 were exposed to the machine-agent cue.

4.2 Experimental Stimulus

Participants were exposed to a screenshot of an iPhone displaying the chat transcript of a user interacting with a chat agent about booking an airline ticket. The last message in the chat sequence asked the user for their credit-card number. The content of the chat was identical in both conditions, except for the source of interaction—either a human sales representative or a machine agent (i.e., Siri; a voice-controlled virtual assistant built in iPhone). Specifically, users in the *human*-agent condition (Figure 1) were shown a human-like icon representing the customer agent, while those in the *machine*-agent condition (Figure 2) were exposed to an interface resembling Siri.



I want to book a one-way flight ticket from New York to Chicago on March 30.

Here is the best ticket option: $114 / from New York (JFK) to Chicago (ORD) / 7:45 am - 9:32 am by US Airways (Non-stop)

I like it, can you book it for me?

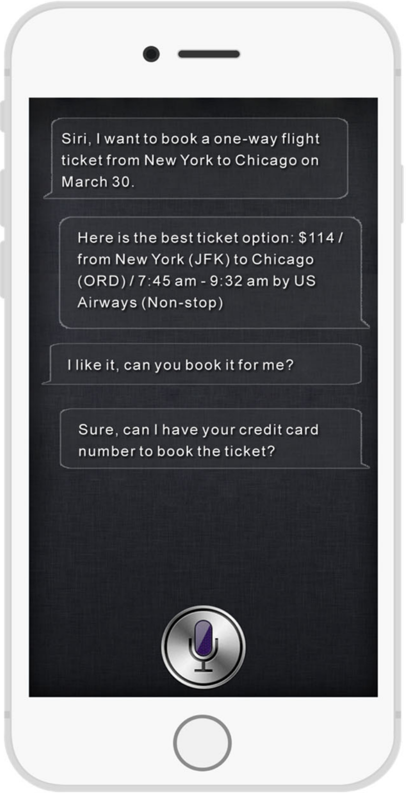
Sure, can I have your credit card number to book the ticket?

Figure 1: The chat session in the Human Agent condition

4.3 Measurements

The following items were used for measuring the dependent variable as well as the moderating, mediating and control variables in the study.

4.3.1 Moderating variable. Belief in machine heuristic was measured via five questionnaire items. Participants answered the degree to which they agreed with each of the following statements about the heuristic on a 7-point Likert scale (1 = strongly disagree; 7 = strongly agree): (1) When machines perform a task, the results are more objective than when humans perform the same task; (2) Machines can handle information in a secure manner, therefore it is okay to disclose my private information to them; (3) Machines have high precision, so they will handle my personal information in a secure way; (4) Machines do not gossip, so they will not share my private information with others; (5) It is safer to reveal personal information to machines rather than to humans. These five items had good reliability (Cronbach's alpha = .87), with a mean of 3.96 and standard deviation of 1.54.



Siri, I want to book a one-way flight ticket from New York to Chicago on March 30.

Here is the best ticket option: $114 / from New York (JFK) to Chicago (ORD) / 7:45 am - 9:32 am by US Airways (Non-stop)

I like it, can you book it for me?

Sure, can I have your credit card number to book the ticket?

Figure 2: The chat session in the Machine Agent condition

4.3.2 Dependent variable. Intention to disclose personal information was measured by a single item [6] on a 7-point scale asking participants their likelihood (1 = very unlikely;

7 = very likely) of entering their credit card number at that juncture of their interaction with the chat agent. The exact wording was "How likely are you to disclose your credit card information to purchase the flight ticket?" (*M* = 3.19, *SD* = 2.25).

4.3.3 Control variables. Previous studies have shown that users' demographic characteristics and other individual differences influence users' disclosure behaviors [24, 28, 31]. Hence, the following variables were measured in the interest of statistically controlling their effects.

4.3.3.1 Demographics. Participants' gender, income, age, race, and education level were measured by the items from the US Census Bureau's 2014 American Community Survey.

4.3.3.2 Privacy concerns. Users' general apprehension of losing their privacy on the Internet was measured with four items from Dinev and Hart [8]. Participants indicated the degree to which they agreed with the perceived risks of losing privacy on a 7-point Likert scale (1 = strongly disagree; 7 = strongly agree). Items were: "I am concerned that the information I submit on the Internet could be misused," "When I shop online, I am concerned that the credit card information could be stolen while being transferred over the Internet," "I am concerned about submitting information on the Internet, because of what others might do with it," and "I am concerned about submitting information on the Internet, because it could be used in a way I did not foresee" (*M* = 5.43, *SD* = 1.23, α = .85).

4.3.3.3 Dispositional distrust. Users' tendency to distrust was assessed with five items from Van Lange et al.'s study [31]. Participants indicated the degree to which they do not trust other individuals on a 7-point Likert scale (1 = strongly disagree; 7 = strongly agree). Items were: "Nowadays you have to be careful, otherwise people will exploit you," "If there were fewer policemen, it would be much more dangerous on the streets," "One should not trust other people, unless one knows them well," "Many things in life often fail because a lot of people pursue their self-interests," and "You have to be careful with strangers until you know you can trust them" (*M* = 5.01, *SD* = 1.07, α = .76).

4.3.3.4 Power usage. Users' self-efficacy and expertise in information technology was measured by twelve items proposed by Marathe, et al. [17]. Participants reported the degree to which they rely on the technology and feel comfortable using it, on a 7-point Likert scale (1 = strongly disagree; 7 = strongly agree). Items were as follows: "I think most of the technological gadgets are complicated to use" (reverse-coded), "I make good use of most of the features available in any technological device," "I have to have the latest version or updates of technological devices (or software) that I use," "Use of information technology has almost replaced my use of paper," "I love exploring all the features that technological gadgets offer," "I often find myself using many technological devices simultaneously," "I prefer to ask friends how to use any new technological gadget instead of trying to figure it out myself" (reverse-coded), "Using any technological device comes easy to me," "I feel like information technology is a part of my daily life," "Using information technology gives me greater control over my work environment," "Using information technology makes it easier to do my work," and" I would feel lost without information technology" (*M* = 4.84, *SD* = .96, α = .81).

4.3.1 Manipulation check. A single manipulation-check item was included at the end of the scenario to ascertain the effectiveness of the experimental manipulation. Participants answered the degree to which they agreed with the statement "I think I had a chat with a machine-based agent to buy a flight ticket" on a 7-point Likert scale (1 = strongly disagree; 7 = strongly agree).

4.4 Procedure

Participants first answered the questions on demographics. They were then directed to two scenarios—the chat scenario and another scenario unrelated to the current study. To control for any order or spillover effects, the order of presenting the scenarios was randomized. After reviewing the scenario description and the screenshot, participants' intentions to disclose their credit-card information in that scenario was measured. After participants finished both scenarios, their individual differences (i.e., general privacy concerns, dispositional distrust, power usership, belief in machine heuristic) were assessed at the end of the questionnaire. When they completed all the questions, they were given a 10-digit code to receive payment.

4.5 Data Analysis

To test the main effect proposed in H1 and the interaction effect proposed in H2, an analysis of covariance using a general linear modeling approach (standard least squares method) was conducted, with the interface cue (machine vs. human agent) as the independent variable, disclosure intention as the dependent variable, belief in the machine heuristic as the moderating variable, and all the control variables as covariates. Data from three participants were excluded in the main analysis because of a large number of missing values.

5 RESULTS

An independent *t*-test revealed that participants in the machine cue condition were significantly more likely to agree that they talked with a machine agent to search for and buy a flight ticket ($M = 6.25$, $SD = .66$) than those in the human cue condition ($M = 1.56$, $SD = .67$), $t(158) = 44.85$, $p < .001$, Cohen's $d = 7.10$. That is, the experimental manipulation was successful.

H1 hypothesized that if Siri asked for users' credit card number, users would be more likely to reveal it than if the same information was requested by a human agent. Results found that the main effect of the machine cue was significant, F (1, 121) = 8.72, $p < .001$, partial $\eta^2 = .07$, showing that users who saw the machine cue in the scenario were more likely to reveal their credit card number (*Least Square Mean* = 4.67, *Std Error* = 2.25) than those exposed to the human cue (*LSM* = 3.63, *SE* = 2.28), supporting H1. Moreover, the main effect of belief in the machine heuristic was significant, F (1, 121) = 17.1, $p < .0001$, partial $\eta^2 = .12$. The higher the belief, the greater the intention to disclose one's credit card information.

H2 proposed greater disclosure intention among those with stronger greater belief in the machine heuristic in the presence of machine, rather than human, agent. This was also supported, by a significant interaction effect between the machine cue and belief in the machine heuristic, F (1, 121) = 6.34, $p < .05$, partial $\eta^2 = .05$. As can be seen in Figure 3, the positive relationship between accessibility of the machine heuristic and disclosure intention was stronger in the presence of machine agent compared to a human agent.

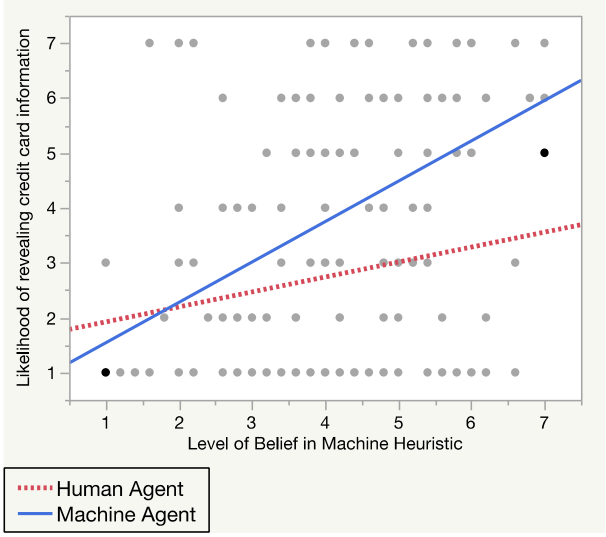


Figure 3: Interaction between Agent Type and Belief in Machine Heuristic upon Disclosure Intention

6 DISCUSSION

Our results suggest, first and foremost, that users trust machine agents over human agents when it comes to disclosing personal information. The higher trust engendered by the machine agent lends support to the notion of automation bias, with users showing greater situational trust in automation [12] in the context of online transactions involving personal financial information.

In addition to showing the effect of interface cues suggesting machine vs. human source of interaction, this study finds that users with greater belief in the trustworthiness of machines tend more to disclose to online agents (whether machine or human) than those with lesser belief in machine trustworthiness. It is noteworthy that we found this effect despite controlling for dispositional trust. It appears that even after accounting for individual differences in the tendency to trust others, there exists an effect due to individual differences in machine trust that predisposes users to engage in sensitive online transactions. That is, regardless of whether the online agent is human or automated, individuals with higher belief in machine heuristic tend to disclose more to online systems, thus suggesting that user interactions are governed by dispositional trust in systems, in addition to the situational trust documented in the automation literature [12].

But, the most important discovery of the current study is the interaction between this dispositional factor and the interface cue of human vs. machine agent of interaction. For individuals with higher levels of belief in the machine heuristic, interface cues that trigger this heuristic led them to greater revelation of their credit card number to Siri. For others who do not believe as strongly in this heuristic, interface cues suggesting machine agency did not increase their intention to reveal such information, to the same degree. Therefore, the effectiveness of interface cues in influencing decision-making appears to be contingent upon the degree of chronic accessibility of the heuristics pertaining to those cues [11].

As Figure 3 clearly shows, even a moderate level of accessibility of the machine heuristic is sufficient for users to detect machine cues on the interface and distinguish them from human cues. Further, consistent with the literature on cognitive accessibility [5, 11], the machine agent led to higher disclosure intention among those who expressed greater belief in the machine heuristic, which substantiates the operation of cognitive heuristics in the decision-making process. If users were simply more impressed by Siri compared to the human agent, then the interaction with prior belief in heuristic would not have been significant. The stronger effect of the machine cue among those with higher belief in the machine heuristic clearly implies the operation of the heuristic in their decision-making about information disclosure. This lends support to the MAIN Model [25], which posits that technological affordances and interface features can influence user judgments by triggering pertinent cognitive heuristics. In this case, the identity of the agent (as human or machine) served as the interface cue that seems to have triggered the machine heuristic. While MAIN model primarily focuses on user perceptions of machines handling news and public information, where objectivity and lack of bias are important, the current study extends this heuristic to the domain of safety and security in handling private information.

By providing the first direct evidence for the operation of the machine heuristic, our data lend empirical support for the heuristic approach to addressing the privacy paradox phenomenon [19]. The study demonstrates the viability of measuring a cognitive heuristic in terms of participants' self-reported agreement with the rule of thumb [5]. More fundamentally, the study's finding speaks to the differential effects of interface cues on users with different belief structures and differential cognitive accessibility of their beliefs.

Several insights for designing more credible and trustworthy interface cues can be derived from this study. First, our discovery of the power of machine-related cue on users' information disclosure suggests the utility of stronger signals of machine agency on interfaces. Given the growing use of artificial intelligence (AI) in computing systems and robots to assist shopping, banking, and other activities, it is important for interfaces to convey to users that the operations are automated and algorithmic in nature, rather than directed by humans. This can engender greater trust, especially in domains where human involvement can lead to unpredictable and/or undesirable outcomes. Additionally, given the significant effect of the machine cue on users' information disclosure as a function of users' belief in the machine heuristic, designers might consider placing a few simple measures in their applications to gauge such differences in trait belief. By doing so, they can better target users who are more (or less) likely to enter their information on the site when shown different types of interface cues, based on individual differences in their level of belief in the machine heuristic.

That said, it is important to stress the ethical use of this heuristic in the design process. The exploitative use of machine heuristic to obtain sensitive information from users (that they would otherwise not provide) is unethical and will erode user trust over time. Such a possibility raises the need for greater digital literacy—not only because a blind belief in machine superiority can be easily exploited by designers, but because machines are not always superior. Users should be not only warned about malevolent use of machine agency (e.g., use of robotic interfaces in phishing attempts), but also made more aware of their susceptibility to automation bias and the persuasive effect of interface cues that suggest machine agency. This could perhaps be achieved with warnings that alert them to their greater susceptibility to suggestions by machine agents compared to human agents. Scholars have suggested the use of "nudges" to assist users in privacy decision-making [2, 4], so nudges that remind users about their relative preference for machines over humans can, at a minimum, help in promoting systematic over heuristic processing of machine cues.

Machine heuristic can have implications that go beyond online interactions with chat agents. It can be invoked in human-robot interaction, for example, by cueing machine, rather than human, aspects of a robot's morphology or interaction. Further, given the growing reliance on automation in almost all domains of activity, this heuristic will likely play a role in decision-makers' trust and reliance in automated solutions for everything from picking stocks and determining recidivism rates to employee selection and retention. Triggering the machine heuristic will promote greater affinity for algorithms in all walks of life, even when users do not quite understand the underlying logic of recommendations made by algorithms. This can have important consequences for both individuals, companies and society, as advances in deep learning and artificial intelligence tend to provide more agency to machines rather than humans.

7 LIMITATIONS AND DIRECTIONS FOR FUTURE RESEARCH

Certain shortcomings of our study design should be addressed by future studies. First, users' intention for information disclosure was assessed in the form of users' self-reports of their willingness to reveal or withhold their private information, after they were asked to imagine that they were in the scenario portrayed in the screenshot. However, there might be a discrepancy between their behavioral intention and actual behavior when faced with a similar situation in reality. Considering the sensitive nature of online interactions with regard to online security and personal privacy, participants might have answered more negatively than their actual tendency to divulge private information on online and mobile sites. Therefore, researchers should consider using experimental stimuli that can capture behavioral data on users' information disclosure instead of asking them to report their intention. Moreover, since individuals tend to prefer easy and expedient decision-making to thorough processing of online information, multiple cognitive heuristics, not a single one, might be at work in the minds of users, such that they can make decisions efficiently. Of course, given the exploratory nature of this study, we only focused on testing the effect of a single heuristic. However, in the scenario where participants were asked to enter their credit card information, they could be influenced not only by the machine heuristic, but also the *authority* heuristic (i.e., popular name, brand, or organization guaranteeing their security) since Siri is a built-in part of iPhone, which is a product of Apple, a well-known brand. Thus, when the participants made their decision, the combined effect of both heuristics could have influenced their behavioral intention. Fortunately, this study examined users' psychological reasons for information disclosure to Siri, including some potential reasons pertaining to the authority heuristic – "Siri is a product of Apple, a company that I trust" (β = -.03, p = .90) and "Apple is a popular company" (β = -.05, p = .79) – but none of these reasons were predictive of their disclosure intentions. Yet, future research could further examine whether a *cue-cumulation effect* [41] occurs in the context of online privacy, increasing users' willingness for disclosure in the presence of multiple interface cues, simultaneously triggering multiple heuristics in the same direction. It is worth noting that trust in the Apple brand may have played a role in our participant's willingness to disclose. However, since the interface was constant across the two conditions, the use of the Apple interface does not pose a confound, but may have inflated disclosure intentions across the board, thus making our study a conservative test. The fact that we found significant differences despite these limitations suggests that the effect of the machine cue is quite robust.

However, statistical significance does not always entail practical significance. Disclosure intentions are generally low, hovering around the mid-point of the scale. This could be because of social desirability bias, since respondents were reporting to study administrators about the extent to which they would be willing to reveal their credit card number online. So, naturally, their estimates are likely to be depressed. Use of more ecologically valid settings and behavioral measures can help overcome this bias and result in higher levels of disclosure, in keeping with actual user behavior.

In conclusion, despite its limitations, our study clearly shows that the machine heuristic can be triggered by an automated agent, with important persuasive consequences. As we interact with newer and more automated entities, user reliance on machine heuristic is likely to increase in the future. Going forth, HCI researchers would also do well to investigate the role of machine heuristic in users' tendency to believe in AI. In particular, it would be useful to ascertain which types of cues on the interfaces of AI-driven technologies, including AI agents and chatbots, are most likely to trigger the machine heuristic. This can not only aid the design of credible virtual agents and trustworthy machines in a number of domains, but also inform the design of digital literacy tools to combat exploitative uses of the machine heuristic.

ACKNOWLEDGMENTS

This research is supported by the U. S. National Science Foundation (NSF) via Standard Grant No. CNS-1450500.

REFERENCES

[1] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein (2015). Privacy and human behavior in the age of information. Science, 509-514. https://doi.org/10.1126/science.aaa1465

[2] Alessandro Acquisti., Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, Florian Schaub, Manya Sleeper, Yang Wang, and Shomir Wilson (2017). Nudges for privacy and security: understanding and assisting users' choices online, ACM Computing Surveys (CSUR) 50 (3), Article No. 44. https://doi.org/10.1145/3054926

[3] Alison Adam (2005). Delegating and distributing morality: Can we inscribe privacy protection in a machine? Ethics and Information Technology, *7*(4), 233-242. https://doi.org/10.1007/s10676-006-0013-3

[4] Hazim Almuhimedi, Florian Schaub, Norman Sadeh, Idris Adjerid, Alessandro Acquisti, J. Gluck, Lorrie Faith Cranor, and Yuvraj Agarwal (2015). Your location has been shared 5,398 times! A field study on mobile app privacy nudging. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI' 15), 787-796. https://doi.org/10.1145/2702123.2702210

[5] Saraswathi Bellur and S. Shyam Sundar (2014). How can we tell when a heuristic has been used? Design and analysis strategies for capturing the operation of heuristics. Communication Methods and Measures, 8(2), 116-137. https://doi.org/10.1080/19312458.2014.903390

[6] Lars Bergkvist and John R. Rossiter (2007). The predictive validity of multiple-item versus single-item measures of the same constructs. Journal of Marketing Research, 44(2), 175-184. https://www.jstor.org/stable/30162466

[7] Erika Chin, Adrienne Porter Felt, Vyas Sekar, and David Wagner (2012). Measuring user confidence in smartphone security and privacy. In Proceedings of the Eighth Symposium on Usable Privacy and Security (SOUPS '12), 1-16. https://doi.org/10.1145/2335356.2335358

[8] Tamara Dinev and Paul Hart (2006). An extended privacy calculus model for e-commerce transactions. Information Systems Research, 17(1), 61- 80. https://doi/10.1287/isre.1060.0080

[9] Susan T. Fiske and Shelley E. Taylor (1991). Social cognition. New York, NY: McGraw-Hill.

[10] Erving Goffman (1967). Interaction ritual: Essays on face-to-face behavior. New York: Anchor Books.

[11] Tory E. Higgins, John A. Bargh, and Wendy J. Lombardi (1985). Nature of priming effects on categorization. Journal of Experimental

Psychology: Learning, Memory, and Cognition, 11(1), 59-69. http://dx.doi.org/10.1037/0278-7393.11.1.59

[12] Kevin Anthony Hoff and Masooda Bashir (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. Human Factors, 57(3), 407-434. https://doi.org/10.1177/0018720814547570

[13] Leslie K. John, Alessandro Acquisti, and George Loewenstein (2011). Strangers on a plane: context-dependent willingness to divulge sensitive information. Journal of Consumer Research, 37(5), 858-873. https://doi.org/10.1086/656423

[14] Vincent Lanaria (2016). Snapchat updates terms of service and privacy policy as chat 2.0 rolls out. Tech Times. Retrieved from http://www.techtimes.com/articles/145613/20160330/snapchat-updates-terms-of-service-and-privacy-policy-as-chat-2-0-rolls-out.htm

[15] Charles Layton, Philip J. Smith, and C. Elaine McCoy (1994). Design of a cooperative problem-solving system for en-route flight planning: An empirical evaluation. Human Factors, 36(1), 94-119. https://doi.org/10.1177/001872089403600106

[16] Gale M. Lucas, Jonathan Gratch, Aisha King, and Louis-Philippe Morency (2014). It’s only a computer: Virtual humans increase willingness to disclose. Computers in Human Behavior, 37, 94-100. https://doi.org/10.1016/j.chb.2014.04.043

[17] Sampada Sameer Marathe, S. Shyam Sundar, Marije Nije Bijvank, Henriette van Vugt, and Jolanda Veldhuis (2007). Who are these power users anyway? Building a psychological profile. Paper presented at the 57th annual conference of the International Communication Association, San Francisco, CA.

[18] Kathleen L. Mosier, Linda J. Skitka, Mark D. Burdick, and Susan T. Heers (1996). Automation bias, accountability, and verification behaviors. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 40(4), 204-208. https://doi.org/10.1177/154193129604000413

[19] Patricia A. Norberg, Daniel R. Horne, and David A. Horne (2007). The privacy paradox: Personal information disclosure intentions versus behaviors. Journal of Consumer Affairs, 41(1), 100-127. https://doi.org/10.1111/j.1745-6606.2006.00070.x

[20] Jonathan A. Obar and Anne Oeldorf-Hirsch (2018). The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. Information, Communication & Society, 7(3), 1-20. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2757465

[21] Raja Parasuraman and Dietrich H. Manzey (2010). Complacency and bias in human use of automation: An attentional integration. Human Factors, 52(3), 381-410. https://doi.org/10.1177/0018720810376055

[22] Matthew D. Pickard, Catherine A. Roster, and Yixing Chen (2016). Revealing sensitive information in personal interviews. Computers in Human Behavior, 65, 23-30. https://doi.org/10.1016/j.chb.2016.08.004

[23] Scott Ruoti, Nathan Kim, Ben Burgon, Timoth van der Horst, and Kent Seamons (2013). Confused Johnny: when automatic encryption leads to confusion and mistakes. In Proceedings of the Ninth Symposium on Usable Privacy and Security (SOUPS ’13*)*, 1-19. https://doi.org/10.1145/2501604.2501609

[24] Pan Shi, Heng Xu, and Yunan Chen, (2013). Using contextual integrity to examine interpersonal-information boundary on social network sites. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’13), 35-38. https://doi.org/10.1145/2470654.2470660

[25] S. Shyam Sundar (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger and A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 72-100). Cambridge, MA: The MIT Press.

[26] S. Shyam Sundar, Hyunjin Kang, Mu Wu, Eun Go, and Bo Zhang (2013). Unlocking the privacy paradox: Do cognitive heuristics hold the key? In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’13*)*, 811-816. https://doi.org/10.1145/2468356.2468501

[27] S. Shyam Sundar, Silvia Knobloch-Westerwick, and Matthias R. Hastall (2007). News cues: Information scent and cognitive heuristics. Journal of the American Society for Information Science and Technology, 58(3), 366-378. https://doi.org/10.1002/asi.20511

[28] S. Shyam Sundar and Sampada Sameer Marathe (2010). Personalization versus customization: The importance of agency, privacy, and power usage. Human Communication Research, 36(3), 298-322. https://doi.org/10.1111/j.1468-2958.2010.01377.x

[29] Amos Tversky and Daniel Kahneman (1974). Judgment under uncertainty: Heuristics and biases. Science, 185(4157), 1124-1131.

[30] Jennifer M. Urban and Chris Jay Hoofnagle (2014). The Privacy Pragmatic as Privacy Vulnerable. SOUPS’14 Workshop on Privacy Personas and Segmentation (PPS). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2514381

[31] Paul A. Van Lange, Mark Van Vugt, Ree M. Meertens, and Rob A. Ruiter (1998). A social dilemma analysis of commuting preferences: The roles of social value orientation and trust. Journal of Applied Social Psychology, 28(9), 796-820. https://doi.org/10.1111/j.1559-1816.1998.tb01732.x

[32] Bo Zhang, Mu Wu, Hyunjin Kang, Eun Go, and S. Shyam Sundar (2014). Effects of security warnings and instant gratification cues on attitudes toward mobile websites. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’14), 111-114. https://doi.org/10.1145/2556288.2557347